

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»  
ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ  
КАФЕДРА МАТЕМАТИЧНИХ МЕТОДІВ СИСТЕМНОГО АНАЛІЗУ

На правах рукопису  
УДК 004.942

До захисту допущено  
В. о. завідувача кафедри ММСА

О.Л.Тимошук

«\_\_» \_\_\_\_\_ 2020 р.

## Магістерська дисертація

на здобуття ступеня магістра за спеціальністю 124 Системний аналіз  
на тему: «Авторегресійні моделі прогнозування навантаження контакт-центру  
компанії та застосування прогнозованих даних»

Виконала:

студентка II курсу, групи КА-91 мп

Діхтяренко Каріна Ігорівна

\_\_\_\_\_

Керівник: доцент кафедри ММСА

к.ф.-м.н, доц. Каніовська І.Ю.

\_\_\_\_\_

Рецензент:

доцент кафедри математичного аналізу

та теорії ймовірностей ФМФ

КПІ ім. Ігоря Сікорського,

к.ф.-м.н, доц. Ільєнко М.К.

\_\_\_\_\_

Засвідчую, що у цій магістерській дисертації  
немає запозичень з праць інших авторів  
без відповідних посилань

Студент \_\_\_\_\_

Київ  
2020

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ  
СІКОРСЬКОГО»  
ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ  
КАФЕДРА МАТЕМАТИЧНИХ МЕТОДІВ СИСТЕМНОГО АНАЛІЗУ

Рівень вищої освіти — другий (магістерський)  
Спеціальність — 124 «Системний аналіз»

ЗАТВЕРДЖУЮ  
В. о. завідувача кафедри ММСА  
О. Л. Тимощук  
«\_\_\_» \_\_\_\_\_ 2020 р.

**ЗАВДАННЯ**  
на магістерську дисертацію студенту Діхтяренко Каріні Ігорівні

**1. Тема дисертації:** «Авторегресійні моделі прогнозування навантаження контакт-центру компанії та застосування прогнозованих даних», науковий керівник дисертації Каніовська Ірина Юріївна, к.ф.-м.н., доцент, затверджені наказом по університету від «02» листопада № 3182-с.

**2. Термін подання студентом дисертації:** 14 грудня 2020 р.

**3. Об'єкт дослідження:** вхідні взаємодії контактного центру, показники ефективності роботи контактного центру.

**4. Предмет дослідження:** авторегресійні моделі, формула Ерланга.

**5. Перелік завдань, які потрібно розробити:**

- 1) Аналіз існуючих методів для прогнозування навантаження.
- 2) Збір, підготовка та аналіз даних для прогнозування вхідних взаємодій.
- 3) Аналіз результатів прогнозу.
- 4) Застосування результатів прогнозу в розрахунку навантаження на контактний центр.
- 5) За результатами розрахунку навантажень підрахунок кількості персоналу для успішної роботи контактного центру.

## 6. Орієнтовний перелік графічного (ілюстративного) матеріалу:

- 1) Структура роботи контактного центру (1.1), методи прогнозування, що застосовуються в контактних центрах (1.2).
- 2) Модель SARIMA (2.1), розподіл Ерланга (2.2), розрахунок навантаження на контактний центр (2.3 – 2.4),
- 3) Вхідні данні (3.1 -3.7), параметри моделей (3.8-3.9), результати прогнозу (3.10-3.12).
- 4) Таблиці з помилками моделей та їх інтерпретації (3.1 -3.3)
- 5) Таблиці у розділі стартап-проекту

## 7. Орієнтовний перелік публікацій:

8. Дата видачі завдання: 01 вересня 2020 р.

### Календарний план

№ з/п	Назва етапів виконання магістерської дисертації	Термін виконання етапів магістерської дисертації
1.	Концептуальний вступ дисертації. Формулювання об'єкта, предмета, цілі, завдань, новизни, практичної значущості результатів	01.09.2020 – 07.09.2020
2.	Перший розділ. Огляд літературно-інформаційних джерел.	08.09.2020 – 28.09.2020
3.	Другий розділ. Дослідження існуючих методів прогнозування, в тому числі авторегресійних моделей.	15.09.2020 – 28.09.2020
4.	Третій розділ. Застосування дослідженого матеріалу на практиці, написання алгоритму обробки великої кількості часових рядів. Аналіз результату прогнозу, написання алгоритму для розрахунку навантаження.	29.09.2020 – 05.10.2020
5.	Четвертий розділ. Стартап-проект	06.10.2020 – 12.10.2020
6.	Концептуальні висновки. Перспективи розвитку отриманих рішень	13.10.2020 – 25.10.2019

Студент

К.І. Діхтяренко

Науковий керівник дисертації

І.Ю. Каніовська

## РЕФЕРАТ

Магістерська дисертація: 75 сторінок, 4 розділи, 18 рисунків, 27 таблиць, 1 додаток, 14 джерел.

**АВТОРЕГРЕСІЯ, КОНТАКТНИЙ ЦЕНТР, ВХІДНІ ВЗАЄМОДІЇ, ПРОГНОЗ, НАВАНТАЖЕННЯ, ФОРМУЛА ЕРЛАНГА.**

Об'єктом дослідження є вхідні взаємодії контактного центру, а також показники ефективності роботи контактного центру.

Предмет дослідження: прогноз вхідних взаємодій контактного центру за допомогою авторегресійних моделей, алгоритм застосування результатів прогнозу для розрахунку навантаження на контактний центр.

Метою даної роботи є розробка робочої моделі для прогнозування вхідних взаємодій, а також алгоритму для автоматичного підбору коефіцієнтів моделі, розробка алгоритму за допомогою якого можна підрахувати необхідну кількість операторів для успішного обслуговування клієнтів компанії.

Отримані результати: результати якісного прогнозу вхідних взаємодій, алгоритм для розрахунку навантаження на контактний центр з урахуванням результатів прогнозу, а також показників ефективності роботи операторів. Програмна реалізація виконана за допомогою мови програмування Python.

Результати роботи були впроваджені в роботу підприємства ТОВ «АВТОДОК УКРЕЙН».

## **ADSTRACT**

Topic: ‘Autoregressive models for predicting contact center load and application of the forecast results’

Thesis: 75 pages, 4 sections, 18 figures, 27 tables, 1 appendix, 14 sources.

AUTOREGRESSION, CONTACT CENTER, INBOUND INTERACTIONS, FORECAST, LOAD, ERLANG'S FORMULA.

The object of the study is the input interactions of the contact center, as well as indicators of the effectiveness of the contact center.

Subject of research: forecast of incoming interactions of the contact center with a help of autoregressive models, algorithm of application the results of the forecast for calculation the contact center load.

The purpose of this work is to develop a working model for predicting incoming interactions, as well as an algorithm for automatic selection of model coefficients, the development of an algorithm that can calculate the required number of operators to successfully serve the company's customers.

The results obtained: the results of a forecast of incoming interactions, an algorithm for calculating the load on the contact center, taking into account the results of the forecast, as well as performance indicators of operators. The software implementation is performed using the Python programming language.

The results of the work were implemented in the work of the company LLC "AUTODOK UKRAINE".

## ЗМІСТ

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ .....	8
ВСТУП .....	9
РОЗДІЛ 1. ДОСЛІДЖЕННЯ ПРЕДМЕТНОЇ ОБЛАСТІ .....	11
1.1 Вступ .....	11
1.2 Визначення контакт-центру та основні канали зв'язку .....	12
1.3 Організація роботи контакт-центру та технічна підтримка .....	14
1.4 Основні показники ефективності контакт-центру .....	15
1.5 Висновки до розділу .....	18
РОЗДІЛ 2. ОПИС МЕТОДІВ ПРОГНОЗУВАННЯ ВХІДНИХ ВЗАЄМОДІЙ ТА СПОСОБУ РОЗРАХУНКУ НАВАНТАЖЕННЯ НА КОНТАКТНИЙ ЦЕНТР .....	19
2.1 Вступ .....	19
2.2 Методи, які використовуються для прогнозування вхідних взаємодій 20	
2.3 Модель ARIMA та її модифікації .....	22
2.4 Розрахунок навантаження на контактний центр .....	28
2.5 Висновки до розділу .....	35
РОЗДІЛ 3. АНАЛІЗ ВХІДНИХ ДАНИХ, ЗАСТОСУВАННЯ ОБРАНИХ МЕТОДІВ ПРОГНОЗУВАННЯ ТА РОЗРАХУНОК НАВАНТАЖЕННЯ НА КОНТНИЙ ЦЕНТР .....	36
3.1 Вступ .....	36

3.2	Аналіз вхідних даних та основних метрик .....	37
3.3	Побудова моделей .....	42
3.3.1	Тест на нормальність .....	42
3.3.2	Вибір параметрів для моделі.....	46
3.3.3	Прогноз.....	47
3.3.4	Аналіз результатів прогнозу .....	49
3.4	Розрахунок навантаження.....	54
3.5	Висновки до розділу.....	55
РОЗДІЛ 4. СТАРТАП-ПРОЕКТ «Contact Center Automation» .....		56
4.1	Вступ.....	56
4.2	Опис ідеї проекту.....	56
4.3	Технологічний аудит ідеї проекту .....	58
4.4	Аналіз ринкових можливостей запуску стартап-проекту .....	59
4.5	Розроблення ринкової стратегії проекту.....	67
4.6	Розроблення маркетингової програми стартап-проекту .....	69
4.7	Висновки до розділу.....	72
ВИСНОВКИ.....		73
ПЕРЕЛІК ПОСИЛАНЬ.....		74

## ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

КЦ – контактний центр.

ARIMA (англ. Auto Regressive Integrated Moving Average) - авторегресійне інтегроване ковзне середнє.

SARIMA (англ. Seasonal Auto Regressive Integrated Moving Average) – сезонне авторегресійне інтегроване ковзне середнє.

AIC (англ. Akaike information criterion) - інформаційний критерій Акаїке.

BIC (англ. Bayesian information criterion) - Баєсів інформаційний критерій.

АНТ (англ. Average Handling Time) - середній час обробки взаємодії.

CSI (англ. Customer Satisfaction Index) – рівень задоволеності клієнта.

FRT (англ. First Response Time) – час першої реакції.

FCR (англ. First Contact Resolution) – успішність вирішення проблеми під час першого дзвінка клієнта.

БД – база даних.

ML (англ. Machine Learning) – машинне навчання.

MAPE (англ. Mean Absolute Percentage Error) – середня абсолютна помилка.

RMSE (англ. Root Mean Squared Error) – корінь середнього квадрату помилки.



## ВСТУП

Контактний центр являється невід’ємною складовою практично будь-якої компанії: від інтернет-магазину до банку. Незважаючи на сучасні автоматизовані методи контакту з клієнтом, залишається велика необхідність у спілкуванні клієнта з операторами сервісу підтримки. Багато хто на власному досвіді переконався, що якість обслуговування контакт-центру напряду впливає на враження клієнта про компанію, а також ймовірність того, що він здійснить покупку, або ж повторно скористається послугами даної компанії. Що ж впливає на якість обслуговування контакт-центру? Не лише правильно підібраний і досвідчений персонал, а ще й правильно розподілене навантаження. Додається складність ще в тому випадку, коли компанія є міжнародною і потрібно обслуговувати клієнтів, що розмовляють різними мовами.

Для налагодженої роботи контакт-центру в провідних успішних компаніях завжди повинен робитися прогноз навантаження на контактний центр для розрахунку виділення бюджету. Оскільки недобір співробітників впливає на якість обслуговування, а перебір – є нераціональним використанням коштів, що впливає на прибуток компанії.

Отже в магістерській дисертації розглядається застосування методів авторегресії для прогнозування кількості вхідних дзвінків контакт-центру міжнародної компанії з продажу автозапчастин і розрахунок навантаження, що в подальшому допоможе дізнатися необхідну кількість працівників, необхідних для налагодженої роботи цього підрозділу компанії, що в свою чергу, допоможе правильно розрахувати бюджет, який необхідно витратити.

Постановка задачі:

1. Аналіз існуючих методів для прогнозування навантаження.
2. Збір, підготовка та аналіз даних для прогнозування вхідних взаємодій.
3. Аналіз результатів прогнозу.
4. Застосування результатів прогнозу для розрахунку навантаження на контактний центр.
5. За результатами розрахунку навантажень підрахунок необхідної кількості персоналу для успішної роботи контактного центру.

## РОЗДІЛ 1. ДОСЛІДЖЕННЯ ПРЕДМЕТНОЇ ОБЛАСТІ

### 1.1 Вступ

Для коректної побудови та аналізу прогнозу, необхідно розуміти, що собою являє контакт-центр і які є специфіки його роботи.

Хоча контакт-центри є технологічно інтенсивними підрозділами, часто 70% і більше їх операційних витрат припадає на людські ресурси, а для мінімізації витрат менеджери ретельно відстежують та прагнуть максимізувати використання оператора. Добре керовані контакт-центри дотримуються чітко визначеного балансу між ефективністю оператора та якістю послуг, і для цього вони використовують теоретичні моделі масового обслуговування. Вхідними даними для цих математичних моделей є статистика щодо системних даних, таких як кількість працюючих операторів, швидкість надходження викликів, час, необхідний для обслуговування клієнта, і тривалість часу, коли клієнти готові чекати на утриманні, перш ніж покладуть слухавку.

Результат – це показники ефективності, такі як розподіл часу, який споживачі очікують "на утриманні", і частка клієнтів, які покидають чергу перед тим, як їх обслуговуватимуть. На практиці кількість операторів, що працюють, стає контрольним параметром, який можна збільшувати або зменшувати для досягнення бажаної компромісної якості.

В даному розділі буде описано загальний процес роботи контакт-центру.

## **1.2 Визначення контакт-центру та основні канали зв'язку**

Контакт-центр – це бізнес-підрозділ організації, який керує взаємодією з клієнтами. На відміну від колл-центру, який отримує запити лише по телефону, контакт-центр обробляє вхідні та вихідні комунікації клієнтів за кількома каналами, такими як телефон, Інтернет, чат, електронна пошта, програми обміну повідомленнями, соціальні медіа, текст, факс та традиційна пошта. Контакт-центри використовують різні типи передових технологій, щоб допомогти швидко вирішити проблеми із клієнтами, відстежувати залучення клієнтів та збирати дані про взаємодію та ефективність. Контакт-центри, як правило, експлуатуються внутрішнім відділом або передаються стороннім постачальникам [1].

В даній роботі буде розглянуто три основних канали зв'язку з клієнтами:

- Дзвінки.

Незалежно від зростаючої кількості доступних опцій, телефонні дзвінки все ще залишаються попереду в галузі контактних центрів. Однак дзвінки, які зараз отримують контактні центри, часто є більш складними, оскільки прості питання дедалі частіше вирішуються за допомогою додатків самообслуговування.

Однією з основних причин збереження популярності телефонних дзвінків є те, що люди часто все ж таки віддають перевагу взаємодії з людьми для вирішення питань. Це спричинено тим, що клієнти вважають, що це заощадить їх час.

Телефон завжди буде важливим каналом. Це також найдорожчий канал обслуговування для контакт-центру.

– Чати.

Не кожен має час або схильність спілкуватися по телефону; це особливо актуально для міленіалів. Альтернативою телефонному дзвінку є веб-чат. Оператори готові надати швидкі відповіді через вікно чату, і це є зручним інструментом для електронної комерції, оскільки надає клієнту можливість продовжувати перегляд, отримуючи швидкі та ефективні відповіді на його запитання.

Веб-чат може бути корисним інструментом для ведення клієнтів через веб-сайт, одночасно маючи змогу зрозуміти, що клієнт розглядає. Зазвичай оператор також може одночасно спілкуватися в чатах з кількома клієнтами, що підвищує продуктивність оператора та зменшує кількість працівників, необхідних для обслуговування робочого навантаження веб-чату.

– E-mail.

Електронна пошта - це найбільш «економний» канал. У середньому вартість обробки електронних листів у п'ять разів нижча, ніж обробка дзвінків. Однак є важливий нюанс: тривалий час реакції.

Електронна пошта є важливим каналом, причому близько третини клієнтів звертаються на електронну пошту як бажаний канал обслуговування клієнтів. Телефонія та електронна пошта складають близько 85% від усіх взаємодій. Клієнтам це зручно, оскільки вони можуть це робити на дозвіллі, і вони не чекають в черзі на відповідь. Але, тим не менш, життєво важливо, щоб на електронні листи реагували вчасно і використовували технології для ефективної обробки електронних листів [2].

### 1.3 Організація роботи контакт-центру та технічна підтримка

Результат даної дипломної роботи було впроваджено в роботу контакт-центру реальної міжнародної компанії.

Отже робота у даному контакт-центрі організована наступним чином:

- Контактний центр складається з близько 1000 співробітників, 3 підрозділів, що базуються у різних країнах (рисунок 1.1).

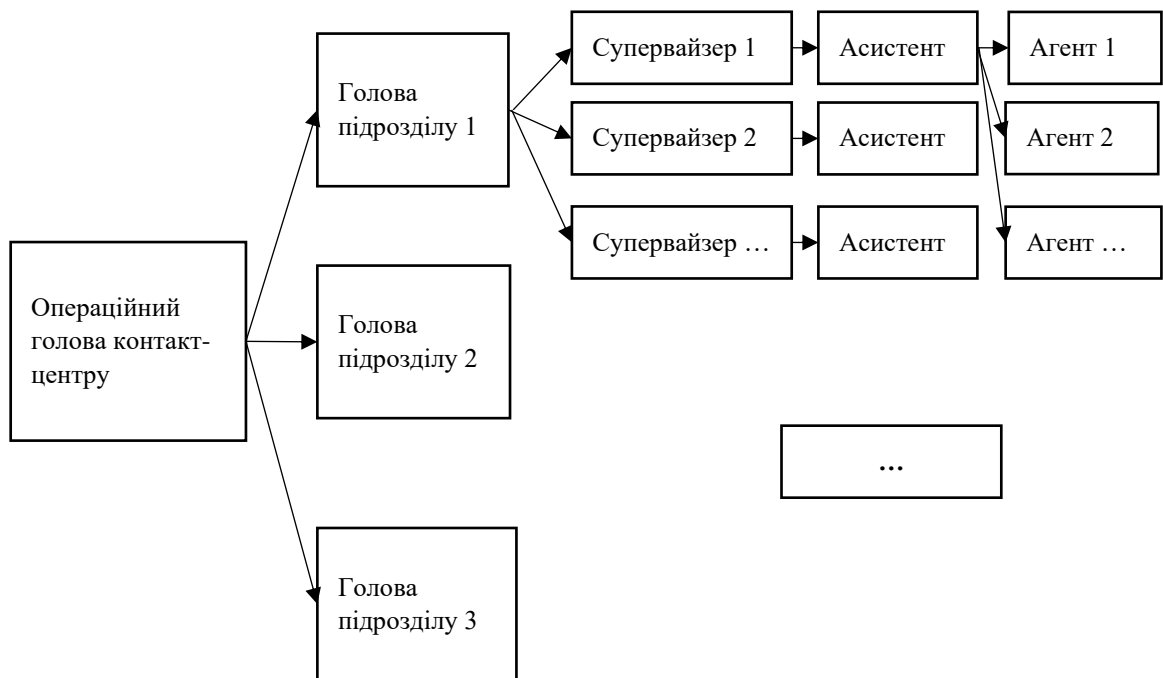


Рисунок 1.1 – Організаційна структура контакт-центру

- Контакт-центр обслуговує низку європейських країн, що розмовляють 22 мовами.

- Система, що використовується для прийому, запису, розподілу дзвінків називається Genesys і представлена однойменною компанією. Вона також записує данні в базу даних на основі MS SQL, звідки і були взяті дані,

що використовуються в даній магістерській дисертації. Ця база даних містить багато детальної технічної інформації про взаємодію, яка відбулася.

- Після того, як клієнт звернувся по одному з трьох вище зазначених каналів зв'язку взаємодія перенаправляється на компетентного оператора, або потрапляє в чергу, де очікує вільного оператора.

- Оператори поділяються на два типи: оператори, які опрацьовують дзвінки та електронні листи і оператори, які задіяні у чатах.

- Якщо мова йде про дзвінки, або електронні листи, оператор не може опрацьовувати більш, ніж 5 взаємодій одночасно, але число може бути меншим і залежить від досвіду і компетентності працівника. Якщо поступає дзвінок, на який може відповісти даний оператор, він стає першим в пріоритеті і в цей момент оператор може опрацьовувати лише цю взаємодію.

- Під час роботи з клієнтом, а також деякий час після цього оператор також працює з CRM-системою, де вносить корективи в замовлення, доповнює, або ж формує його.

- Черга в розумінні КЦ – це своєрідний стек, с якому знаходяться взаємодії до того, як розподілитися на вільного оператора.

- Агент – часто вживане в контактному центрі слово, в контексті КЦ є синонімом до слова «оператор».

#### **1.4 Основні показники ефективності контакт-центру**

Є низка основних показників, завдяки яким вимірюється ефективність роботи КЦ і виявляються слабкі місця. Всі ці показники безпосередньо пов'язані з кількістю вхідних дзвінків, а, отже і з завантаженістю контактного центру [4].

– Customer Satisfaction Index (CSI) – рівень задоволеності клієнта сервісом обслуговування. Зазвичай клієнту пропонується оцінити рівень обслуговування, наприклад, від 1 до 5, після чого ці оцінки усереднюють в різних розрізах.

– First Response Time (FRT) – час першої реакції. Розраховується за формулою:

$$FRT = \frac{\text{Загальний час очікування на всі дзвінки}}{\text{Загальна кількість дзвінків}}$$

– First Contact Resolution (FCR) – відображає успішність вирішення проблеми під час першого дзвінка клієнта. Швидке вирішення проблеми без зворотного дзвінка клієнта сприяє чудовій взаємодії з клієнтами. Розраховується за формулою:

$$FCR = \frac{\text{Загальна кількість взаємодій, вирішених під час першого дзвінка}}{\text{Загальна кількість взаємодій}}$$

– Average Handle Time (АНТ) – середня кількість часу на обробку взаємодії від початку до кінця. Розраховується за формулою:

$$АНТ = \frac{\text{Загальний час розмови} + \text{Загальний час утримання} + \text{Загальний час роботи після взаємодії}}{\text{Загальна кількість взаємодій}}$$

– Abandonment Rate – рівень відмови від дзвінків. Розраховується за формулою:



$$\text{Abandonment Rate} = \frac{\text{Кількість дзвінків} - \text{Кількість оброблених дзвінків}}{\text{Загальна кількість дзвінків}}$$

– Average Waiting Time – середній час очікування. Розраховується за формулою:

$$\text{Average Waiting Time} = \frac{\text{Загальна кількість секунд, які споживачі проводять в очікуванні}}{\text{Загальна кількість дзвінків}}$$

– Service Level - рівень обслуговування. Цей показник служить надійним показником загальної ефективності роботи контакт-центру. Формули:

$$\text{Service Level} = \frac{\text{Кількість дзвінків, прийнятих за } Y \text{ секунд}}{\text{Загальна кількість отриманих дзвінків}} \times 100\%$$

$$\text{Service Level} = \frac{\text{Кількість дзвінків, прийнятих за } Y \text{ секунд}}{\text{Загальна кількість дзвінків, на які відповіли} + \text{Кількість покинутих дзвінків}} \times 100\%$$

$$\text{Service Level} = \frac{\text{Кількість дзвінків, прийнятих за } Y \text{ секунд}}{\text{Загальна кількість дзвінків, на які відповіли} + \text{Кількість покинутих дзвінків за } Y \text{ секунд}} \times 100\%$$

– Occupancy Rate – «зайнятість» агента. Вимірює час, який агенти контакт-центру витрачають на обробку дзвінків у реальному часі та виконання адміністративних завдань, пов'язаних із цими дзвінками. Формула:

$$\text{Occupancy Rate} = \frac{\text{Загальний час обробки контактів}}{\text{Загальний час в системі}} \times 100\%$$

– Shrinkage – показує різницю між часом, який агенти витрачають на відповідь на дзвінки та обслуговування своїх клієнтів, та їх фактичний робочий час.

### **1.5 Висновки до розділу**

В даному розділі було описано загальний принцип роботи контакт-центру, основні метрики, які використовуються для оцінки його ефективності. Всі поняття, описані вище, є важливою складовою для розрахунку навантаження КЦ.

## **РОЗДІЛ 2. ОПИС МЕТОДІВ ПРОГНОЗУВАННЯ ВХІДНИХ ВЗАЄМОДІЙ ТА СПОСОБУ РОЗРАХУНКУ НАВАНТАЖЕННЯ НА КОНТАКТНИЙ ЦЕНТР**

### **2.1 Вступ**

Як в контактному центрі, так і в банківському відділенні, менеджери скрізь стикаються з постійною проблемою збалансування пріоритетів рівня послуг та витрат на оплату праці. Це особливо актуально в сучасних економічних умовах, коли неточні прогнози можуть призвести до неправильних рішень щодо планування та можуть мати значний вплив на зниження прибутковості.

В цьому розділі буде описано теоретичну частину методів, які використовуватимуться для прогнозування навантаження на контактний центр.

В моєму випадку найбільшою складністю є те, що через масштаби компанії (велику кількість країн, що розмовляють різними мовами), а також 3 каналів взаємодій (дзвінки, чати та електронні листи), маємо завелику кількість часових рядів (а саме 66) для того, щоб детально аналізувати кожен, а, отже, потрібно розглядати ті моделі, де можна застосувати автоматичний підбір коефіцієнтів. Також не буде можливості провести детальний візуальний аналіз даних.

## 2.2 Методи, які використовуються для прогнозування вхідних взаємодій

Станом на 2016 рік майже половина контактних центрів використовує «ручні» методи прогнозування навантаження на контактний центр (наприклад, за допомогою функцій Excel). До того, як я приступила до цієї задачі, так було і в даному КЦ.

На графіку нижче (рисунок 2.1) зображено розподіл методів прогнозування вхідних взаємодій, які використовуються різними КЦ [5]:

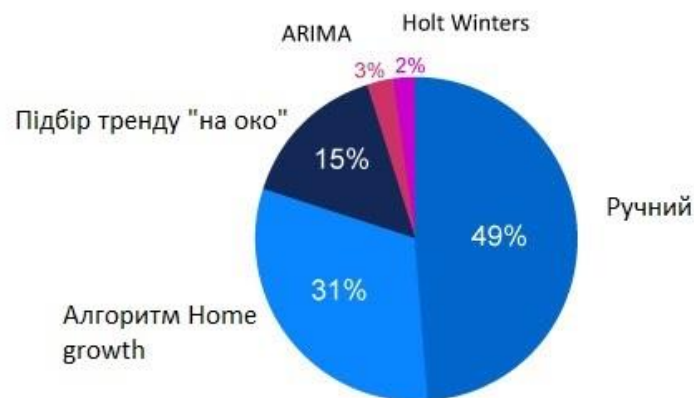


Рисунок 2.1 – Розподіл методів прогнозування, що використовуються КЦ

З рисунку видно, що на практиці, лише маленька частка контактних центрів, використовують більш складні алгоритми прогнозування даних.

Згідно з інтернет-ресурсами, на даний момент актуальними є наступні 4 підходи до прогнозування:

- Потрійне експоненціальне згладжування.

Цей метод передбачає розподіл даних на три компоненти: рівень, тренд та сезонність.

– ARIMA (Auto Regressive Integrated Moving Average) – авторегресійне інтегроване ковзне середнє.

Цей метод теж складається з трьох основних компонент:

– Авторегресія – можливість порівняння даних із попередніми шаблонами.

– Інтегрований – можливість порівняння поточного спостереження від попереднього спостереження.

– Ковзне середнє – здатність згладжувати дані за низку минулих періодів.

– Нейронні мережі.

– MTA (Multiple Temporal Aggregation) – множинна тимчасова агрегація.

Ця стратегія поєднує погодинні, щоденні, щотижневі дані з довгостроковими тенденціями. Це дозволяє контактному центру зосередитись на щоденному прогнозуванні, заглядаючи наперед.

Враховуючи кількість часових рядів, при цьому не дуже великий часовий проміжок з 2017 року з щомісячними даними (що не дуже підходить для нейронних мереж), я вирішила використовувати алгоритм ARIMA та його модифікації. Прогноз на щомісячній основі (а не частіше) спричинений тим, що для планування набору персоналу така частота є оптимальною.

## 2.3 Модель ARIMA та її модифікації

Експоненціальне згладжування та моделі ARIMA є двома найбільш широко використовуваними підходами до прогнозування часових рядів.

У моделях множинної регресії ми прогнозуємо змінну, використовуючи лінійну комбінацію предикторів. В моделі авторегресії ми прогнозуємо змінну, використовуючи лінійну комбінацію минулих значень змінної. Термін авторегресія вказує на те, що це регресія змінної щодо себе.

Таким чином, модель авторегресії порядку  $p$  може бути записана як:

$$y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + \varepsilon_t,$$

де  $\varepsilon_t$  – білий шум.

Ми називаємо це моделлю  $AR(p)$ , моделлю авторегресії порядку  $p$ .

Авторегресійні моделі надзвичайно гнучкі при роботі з широким спектром різних моделей часових рядів.

Замість того, щоб використовувати минулі значення змінної прогнозу в регресії, модель ковзного середнього використовує минулі помилки прогнозу в моделі, подібній до регресії.

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_p \varepsilon_{t-p},$$

де  $\varepsilon_t$  – білий шум.

Це називається моделлю МА ( $q$ ), моделлю ковзного середнього порядку  $q$ .

Ми не спостерігаємо значень  $\varepsilon_t$ , тому насправді це не регресія у звичному розумінні.

Модель ковзного середнього використовується для прогнозування майбутніх значень, тоді як згладжування ковзного середнього використовується для оцінки циклу тренду минулих значень.

Якщо ми поєднуємо авторегресію та модель ковзного середнього, ми отримуємо несезонну модель ARIMA. ARIMA є аббревіатурою від Autoregressive Integrated Moving Average (у цьому контексті "інтеграція" є зворотною стороною різниці).

ARIMA можна записати як:

$$y'_t = c + \varphi_1 y'_{t-1} + \varphi_2 y'_{t-2} + \dots + \varphi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

де  $y'_t$  – різницева серія.

До «предикторів» з правого боку належать як значення  $y_t$  з лагом, так і помилки з лагом.

Це називається моделлю  $ARIMA(p, d, q)$ , де:

- $p$  – порядок частини з авторегресією;
- $d$  – ступінь першого диференціювання;
- $q$  – порядок частини ковзного середнього.

Як тільки ми починаємо поєднувати компоненти таким чином, щоб сформувати більш складні моделі, набагато простіше працювати з позначенням зворотного зсуву. Наприклад, рівняння можна записати як:

$$(1 - \varphi_1 B - \dots - \varphi_p B^p)(1 - B)^d y_t = c + (1 + \theta_1 B + \dots + \theta_q B^q) \varepsilon_t$$

Ті самі умови стаціонарності та зворотності, що використовуються для моделей авторегресії та ковзного середнього, також застосовуються до моделі ARIMA.

Після того, як був визначений порядок моделі (тобто значення  $p$ ,  $d$  і  $q$ ), нам потрібно оцінити параметри  $c, \varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q$ . Це можна зробити, наприклад, за допомогою методу максимальної правдоподібності (MLE).

Цей метод знаходить значення параметрів, які максимізують ймовірність отримання даних, які ми спостерігали.

Для моделей ARIMA показник MLE подібний до оцінок найменших квадратів, які можна отримати при мінімізації  $\sum_{t=1}^T \varepsilon_t^2$ .

Інформаційний критерій Акаїке (AIC), який був корисний при виборі предикторів для регресії, також корисний для визначення порядку моделі ARIMA. Це можна записати як

$$AIC = -2 \log(L) + 2(p + q + k + 1),$$

де  $L$  - ймовірність даних;

$k = 1$ , якщо  $c \neq 0$  і  $k = 0$ , якщо  $c = 0$ .

Останній доданок у дужках - це кількість параметрів моделі (включаючи  $\sigma^2$ , дисперсію залишків).

Для моделей ARIMA виправлений AIC можна записати як:

$$AIC_c = AIC + \frac{2(p + q + k + 1)(p + q + k + 2)}{T - p - q - k - 2},$$

а інформаційний критерій Баєса можна записати як:

$$BIC = AIC + [\log(T) - 2](p + q + k + 1)$$





$$\varphi_p(L)\tilde{\varphi}_P(L^S)\Delta^d\Delta_S^D y_t = A(t) + \theta_q(L)\tilde{\theta}_Q(L^S)\varepsilon_t,$$

де  $\varphi_p(L)$  – несезонний поліном авторегресії з лагом;

$\tilde{\varphi}_P(L^S)$  – сезонний поліном авторегресії з лагом;

$\Delta^d\Delta_S^D y_t$  – часовий ряд з різницею  $d$  і сезонною різницею  $D$ ;

$A(t)$  – поліном тренду;

$\theta_q(L)$  – несезонний поліном ковзного середнього з лагом;

$\tilde{\theta}_Q(L^S)$  – сезонний поліном ковзного середнього з лагом.

Сезонна частина моделі складається з термінів, подібних до несезонних компонентів моделі, але передбачає зворотний зсув сезонного періоду. Наприклад,  $ARIMA(1,1,1)(1,1,1)_4$  призначена для квартальних даних ( $m = 4$ ) і може бути записана як:

$$(1 - \varphi_1 B)(1 - \varphi_1 B^4)(1 - B)(1 - B^4)y_t = (1 + \theta_1 B)(1 + \theta_1 B^4)\varepsilon_t$$

Додаткові сезонні умови просто перемножуються на несезонні.

Тепер розглянемо ще одну модифікацію SARIMAX.

Ця модель демонструє використання “пояснювальних” змінних (Х частина ARIMAX). Коли включені екзогенні регресори, модель SARIMAX використовує поняття «регресія з помилками SARIMA і записується наступним чином:

$$y_t = \beta_t x_t + u_t$$

$$\varphi_p(L)\tilde{\varphi}_P(L^S)\Delta^d\Delta_S^D u_t = A(t) + \theta_q(L)\tilde{\theta}_Q(L^S)\varepsilon_t$$

Варто зауважити, що перше рівняння – це лінійна регресія, а друге рівняння описує процес, за яким слідує компонент помилки. Однією з причин

цією специфікації є те, що оцінювані параметри мають свої природні інтерпретації.

## 2.4 Розрахунок навантаження на контактний центр

Великий вклад в розрахунок навантаження на контактні центри зробив Ерланг.

Спочатку поговоримо про розподіл Ерланга.

Розподіл Ерланга (який іноді називають розподілом Ерланга- $k$ ) був розроблений А.К. Ерлангом, щоб знайти кількість телефонних дзвінків, які можна здійснити одночасно операторам комутаційних станцій. Ерланг був інженером з телекомунікацій у телефонній компанії Копенгагену. Його формули збитків та часу очікування використовували багато телефонних компаній, включаючи Британське поштове відділення. З тих пір розподіл Ерланга було розширено для використання в теорії черг, математичному вивченні черги. Він також використовується в стохастичних процесах і в математичній біології [7].

Розподіл Ерланга є особливим випадком гамма-розподілу (рисунок 2.2). Він визначається двома параметрами,  $k$  і  $\mu$ , де:

- $k$  - параметр фігури. Це має бути натуральне ціле число. У розподілі гамми  $k$  може бути будь-яке дійсне число.
- $\mu$  - параметр шкали. Має бути додатним дійсним числом.

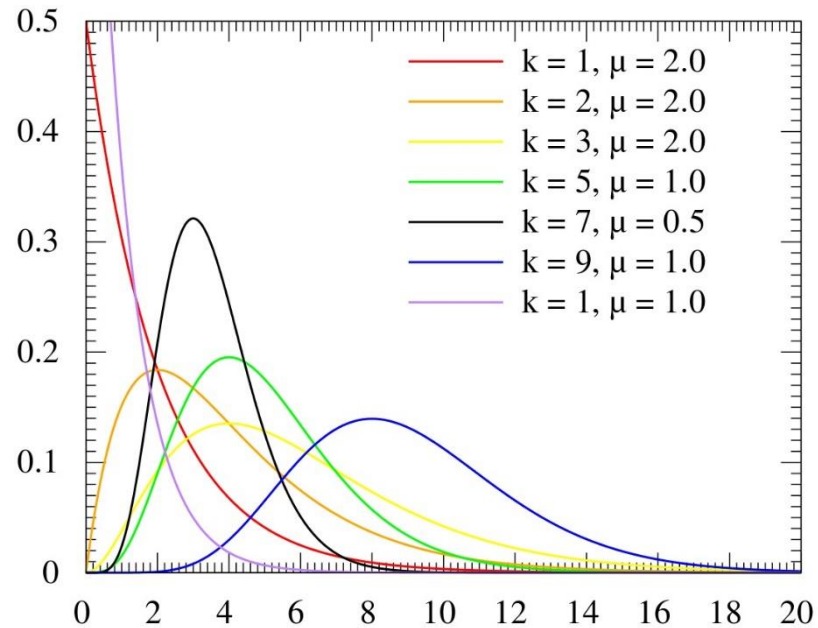


Рисунок 2.2 – Різні можливі параметри форми та масштабу розподілу Ерланга

Функція розподілу ймовірності має наступний вигляд:

$$f(x; k; \mu) = \frac{x^{k-1} e^{-\frac{x}{\mu}}}{\mu^k (k-1)!},$$

де  $k, \mu \geq 0$ .

Факторіал у знаменнику зумовлений тим, що розподіл визначений лише для додатних чисел. Еквівалентна форма для цього розподілу включає  $\lambda$ , міру швидкості, яка пов'язана з  $\mu$  наступним чином:  $\mu = \frac{1}{\lambda}$ .  $\lambda$  являє собою кількість елементів або дзвінків, очікуваних за певний проміжок часу.

Числові характеристики:

- Математичне сподівання:  $\frac{k}{\lambda}$ .
- Дисперсія:  $\frac{k}{\lambda^2}$ .

Зв'язок з іншими розподілами:

- Гамма-розподіл є узагальненою формою розподілу Ерланга.
- Якщо параметр  $k$  дорівнює 1, розподіл такий самий, як експоненційний.
- Якщо параметр масштабу дорівнює 2, розподіл такий самий, як  $\chi^2$ -квадрат з 2 ступенями свободи.

Тепер ознайомимося з формулами Ерланга.

Три формули Ерланга були розроблені в першій половині 20 століття: Ерланг А, Ерланг В і Ерланг С [9].

Однак ці формули були створені не в такому порядку, причому Ерланг В була першою, яку винайшов датський математик на ім'я А. К. Ерланг.

У 1917 році Агнер Крауп Ерланг продовжив свої дослідження, створивши формулу Ерланг С, перш ніж інший математик, на ім'я Конні Палм, винайшов формулу Ерланг А в 1946 році.

Хоча формула Ерланг В сьогодні насправді не використовується в контактних центрах, тоді як Ерланг С і Ерланг А все ще мають величезну роль в управлінні робочою силою контактного центру.

Спочатку формула Ерланг В була розроблена для того, щоб визначити ймовірність блокування в телефонній системі. Отже, коли контактні центри були вперше введені, Ерланг В використовувався для розрахунку, скільки ліній вам потрібно в контактному центрі, щоб підтримувати чергу.

Формула Ерланга В має наступний вигляд:

$$Pb = \frac{\frac{A^N}{N!}}{\sum_{X=0}^N \frac{A^X}{N}},$$

де  $A$  – вхідний трафік;  
 $N$  – кількість ліній;  
 $Pb$  – ймовірність «відмови».

Отже, вхідні дані в Ерланг В при застосуванні до контактному центру включатимуть прогнозовані обсяги дзвінків та середній час обробки (АНТ) для обчислення середньої швидкості переміщення черги. Ці значення впливають на необхідну кількість телефонних ліній.

Формула Ерланг С, побудована на формулі Ерланг В, використовується в контактному центрі для визначення кількості операторів, яка необхідна в контактному центрі для досягнення заданого рівня обслуговування.

Формула Ерланг С має наступний вигляд:

$$P = \frac{\frac{A^N}{N!} \left( \frac{N}{N-A} \right)}{\sum_{X=0}^{N-1} \frac{A^X}{X!} + \frac{A^N}{N!} \left( \frac{N}{N-A} \right)},$$

Формула кодується у пристрої, відомому як "Калькулятор Ерланга", що полегшує розрахунки персоналу контактному центру.

Для цього Калькулятор Ерланга перетворює кількість дзвінків, АНТ та рівень обслуговування у кількість необхідного персоналу, тоді як сучасні версії також враховують інші показники – такі як заповнення, усадка та середнє терпіння.

Однак для розрахунку середнього терпіння також знадобиться невелика допомога від формули Ерланга А.

Формула Erlang C, хоча і була дуже корисною, не враховувала кількість людей, які відмовляються від дзвінків (кидають трубку), перш ніж звернутися до консультанта.

На щастя, у 1946 р. Шведський математик Конні Палм розробив розширення формули Ерланг С, яка відома як формула Ерланга А (рисунок 2.3), яка дійсно враховує "відмову". Насправді А в Ерланг А означає «Відмовляються» (Abandoned).

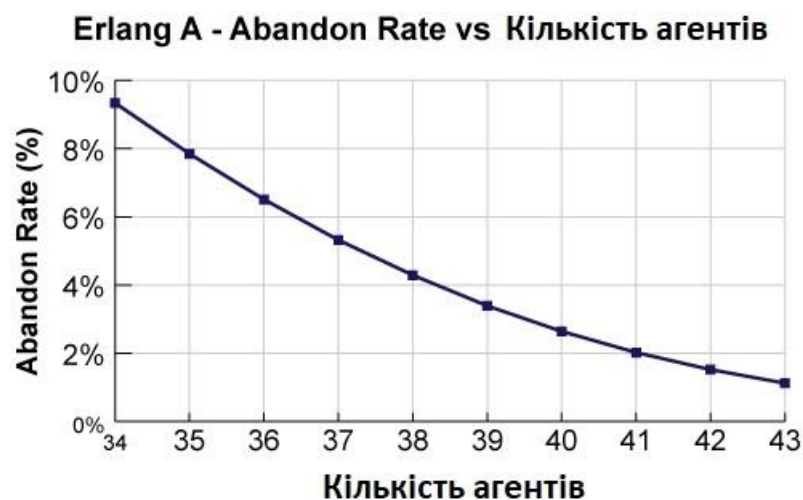


Рисунок 2.3 – Залежність кількості операторів від кількості відмов, знайдена за допомогою формули Ерланга А

Створюючи формулу Ерланга А, Палм використав простий, керований спосіб вимірювання відмов, який поєднує математику А. К. Ерланга та Андрія Маркова.

Ерланг А впливає з теорій Маркова, а також Ерланга В & С.

Формула Ерланга А, будучи продовженням Ерланга В і С, враховує процес «народження й смерть», теоретизована російським математиком Андрієм Марковим.



Потім Конні Палм це інтерпретував у своїй формулі Ерланга А, де "народження" - це люди, які входять в чергу, а "смерть" - це люди, які відмовляються від черги, що робить теорію Маркова основою формули.

Об'єднавши математику А. К. Ерланга та Андрія Маркова, рівняння Ерланга А. Палма виявилось чудовим методом прогнозування «відмови» абонента.

Формула Ерланга А має наступний вигляд:

$$P\left\{\frac{Ab}{W} > 0\right\} = \frac{1}{\rho A\left(\frac{\eta\mu}{\theta}, \frac{\lambda}{\theta}\right)} + 1 - \frac{1}{\rho}$$

Основна відмінність полягає в тому, що крім рівня обслуговування, також потрібно враховувати середнє терпіння.

Середнє терпіння - це швидкість, з якою 50% людей відмовляться, яку і яку досить легко розрахувати у контактному центрі, побудувавши відсоток відмовлених дзвінків від часу (рисунок 2.4).

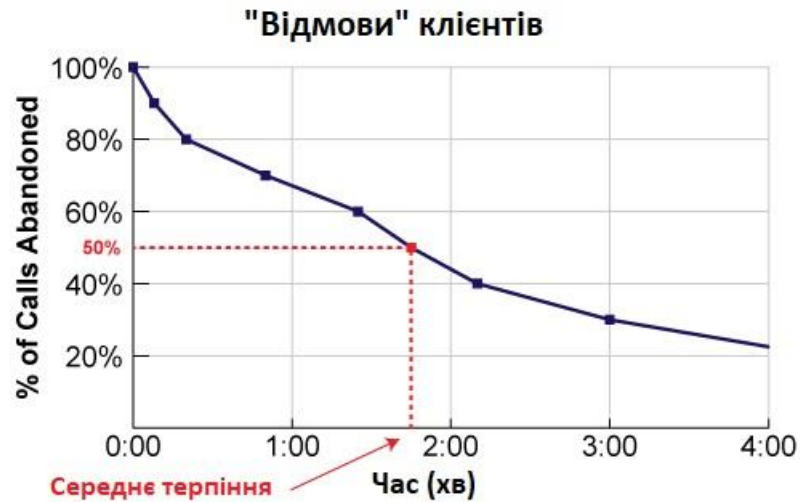


Рисунок 2.4 – Приклад візуального відображення «середнього терпіння»

Незважаючи на те, що Ерланг А добре оцінює кількість «відмовлених» дзвінків на основі «середнього терпіння», він має тенденцію недооцінювати кількість необхідного персоналу.

Наприклад, якщо зателефонувало 50 людей, а Ерланг А передбачає, що 10 відмовляться, то, якщо просто скористатися формулою Ерланг А, то результатом буде лише 40 працівників. Однак будуть деякі дні, коли менше або більше 10 людей «кидають трубку». Через це важливо не використовувати формулу Ерланга А поодинці, кількості людей для контактного центру.

Отже, поєднуючи Ерланга С та Ерланга А, можна отримати найбільш точний прогноз кількості необхідного персоналу відповідно до наступних показників:

- прогнозовані обсяги вхідних взаємодій;
- середній час обробки (АНТ);
- рівень обслуговування (Service Level);
- заповнюваність;
- зайнятість (Occupancy Rate);
- середнє терпіння.

## 2.5 Висновки до розділу

В цьому розділі було описано існуючі підходи до прогнозування вхідних взаємодій контактного центру, в тому числі ті, що будуть розглядатися надалі: ARIMA, SARIMA, SARIMAX; та формули для розрахунку навантаження на КЦ, де застосовується цей прогноз: Ерланга А, Ерланга В, Ерланга С.

### **РОЗДІЛ 3. АНАЛІЗ ВХІДНИХ ДАНИХ, ЗАСТОСУВАННЯ ОБРАНИХ МЕТОДІВ ПРОГНОЗУВАННЯ ТА РОЗРАХУНОК НАВАНТАЖЕННЯ НА КОНТНИЙ ЦЕНТР**

#### **3.1 Вступ**

В даному розділі буде описано особливості вхідних даних, обґрунтовано вибір моделей для прогнозування та зроблено аналіз отриманих результатів.

Всі вхідні дані було взято з бази даних на основі Microsoft SQL. Дані в цій базі являють собою записи з детальною інформацією про кожну взаємодію: унікальний ідентифікатор, час взаємодії, вид взаємодії, канал взаємодії (чат / електронний лист / дзвінок), черга (мова оператора, яким буде прийнята взаємодія, в нашому випадку є 22 мови) та багато іншої інформації. За допомогою цієї інформації, можна звести дані до необхідного нам вигляду: щомісячні зрізи кількості вхідних взаємодій по кожному каналу та всі необхідні показники ефективності (1.1 - 1.8), що описані в розділі 1.4.

Обробка, аналіз, моделювання, та розрахунок навантаження виконано за допомогою мови програмування Python у веб-орієнтованому інтерактивному середовищі Jupyter Notebook. Для цього знадобилися наступні (основні) бібліотеки:

- pyodbc – для звернення до серверу БД безпосередньо за допомогою SQL-запитів;
- pandas – для роботи з таблицями (так званими “data frames”);
- matplotlib – для візуалізації даних;
- numpy – для роботи з масивами даних;
- statsmodels – для моделювання;

– scіру – для аналізу даних.

### **3.2 Аналіз вхідних даних та основних метрик**

Для якісного прогнозу часових рядів необхідна якомога більша кількість історичних даних, саме тому я взяла максимально можливий об'єм інформації, який зберігається в наші базі даних. Отже, маємо наступну доступність даних: дзвінки і чати – з грудня 2017 року, а електронні листи – з вересня 2018 року. З урахуванням того, що необхідно зробити прогноз на 12 місяців вперед (для планування бюджету контактного центру), на жаль, маємо доволі малий історичний період, що вплине на якість прогнозу.

Оскільки на вхід маємо 66 часових ряди, ми не маємо можливості детально проаналізувати кожен із них, але все ж без попереднього аналізу даних неможливий якісний прогноз.

Отже, розглянемо загальну тенденцію часових рядів.

Вхідні дзвінки мають наступну динаміку (рисунок 3.1):

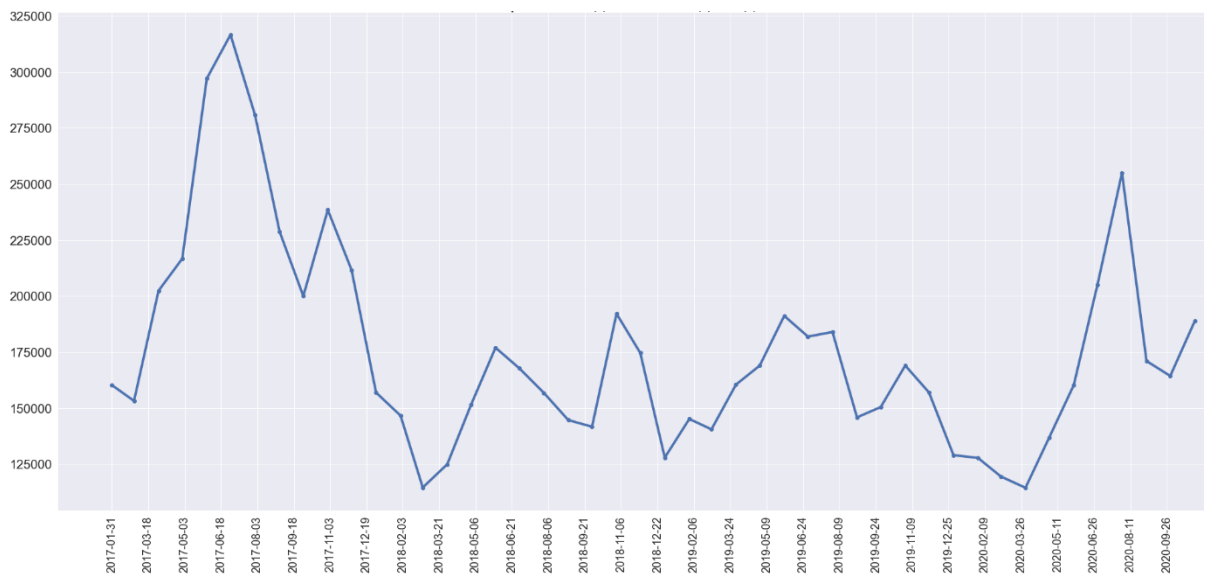


Рисунок 3.1 – Щомісячна динаміка кількості вхідних дзвінків

Незважаючи на те, що піку по кількості вхідні дзвінки досягли в 2017 році, можна помітити, що за останні півроку динаміка починає знову рости. Це пов'язано з запуском маркетингових кампаній влітку цього року.

Чати мають наступну динаміку (рисунок 3.2):

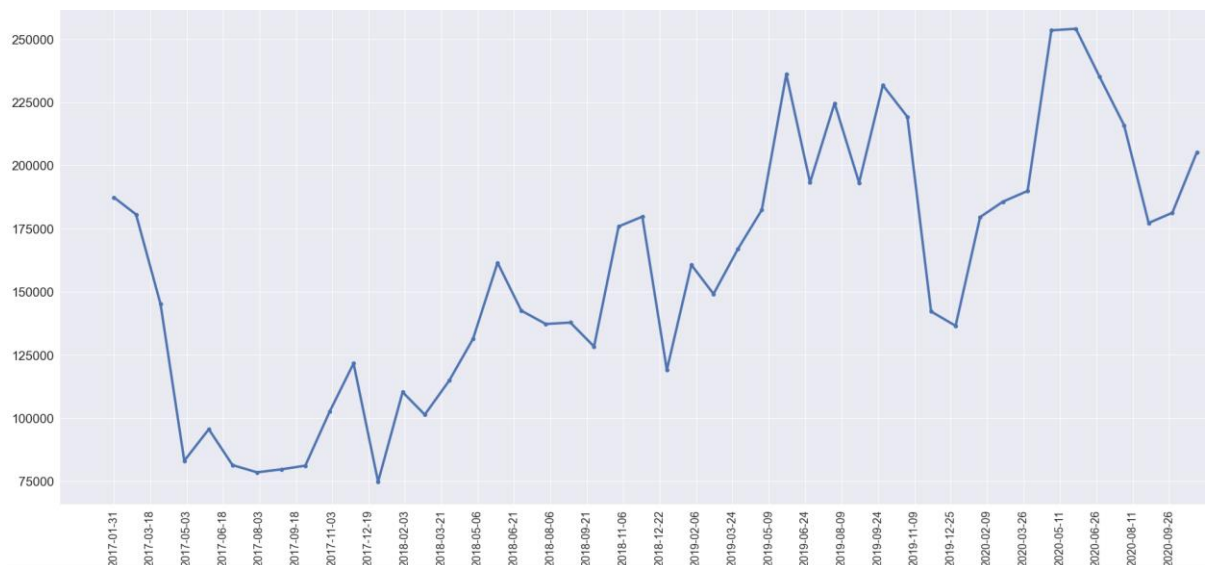


Рисунок 3.2 – Щомісячна динаміка кількості чатів

На відміну від дзвінків, чати мають яскраво виражену тенденцію росту із сезонними явищами.

Електронні листи мають наступну динаміку (рисунк 3.3):

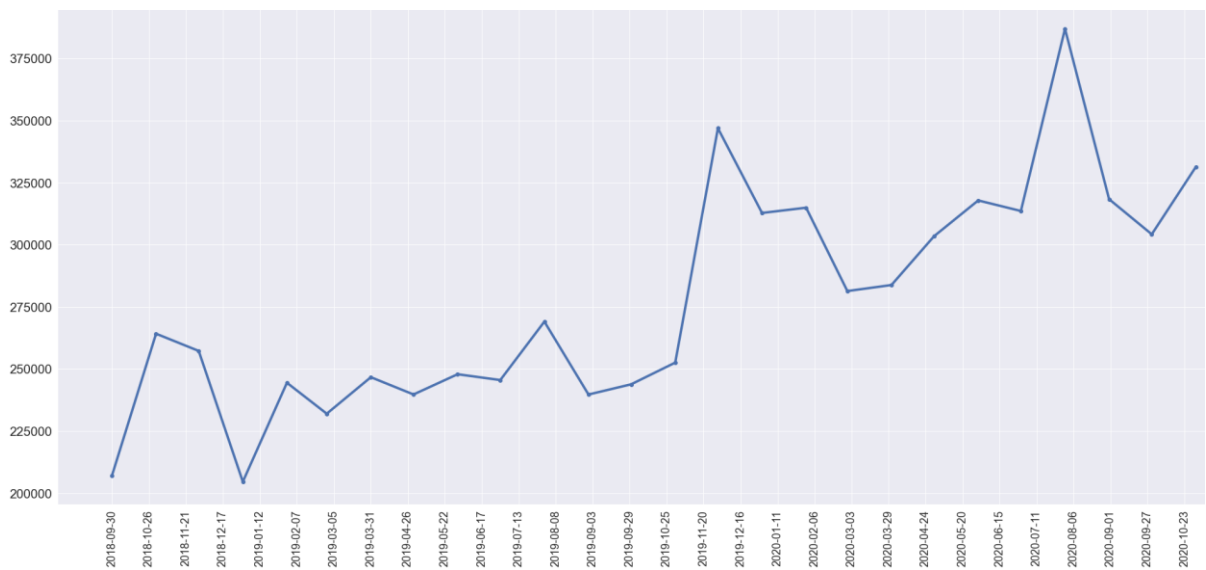


Рисунок 3.3 – Щомісячна динаміка кількості вхідних електронних листів

Електронні листи – найпопулярніший вид взаємодій в даній компанії і, аналогічно до чатів, має тенденцію росту.

Подивимося на динаміку вхідних дзвінків в розрізі 7 найпопулярніших (найбільших за об'ємом) каналів зв'язку (рисунк 3.4).

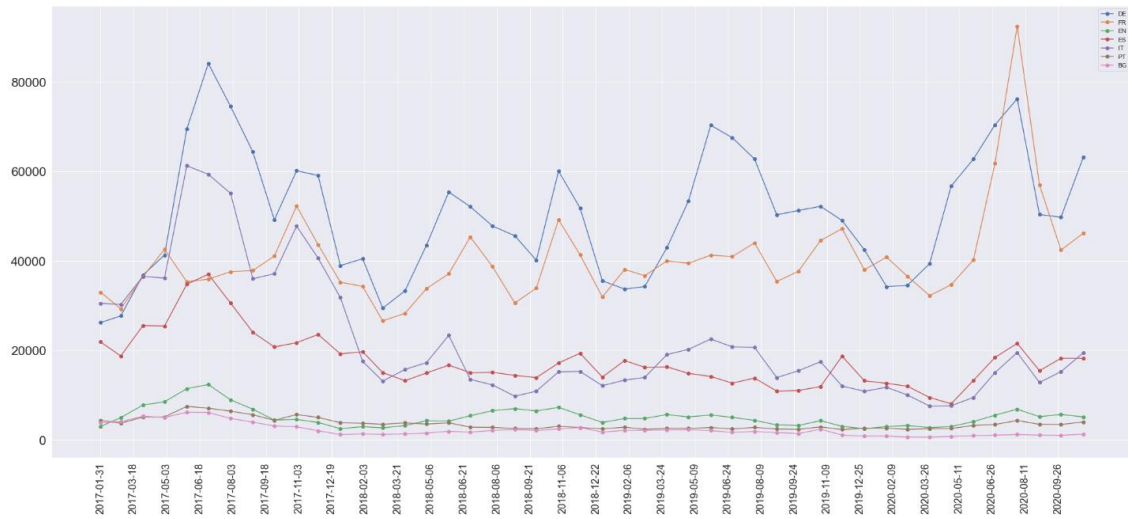


Рисунок 3.4 – Щомісячна динаміка кількості вхідних дзвінків за сімома основними каналами зв'язку

На цьому графіку можна побачити, що тренди і піки по різних каналах зв'язку є доволі закономірними. Тому вже на цьому етапі аналізу вхідних даних можна схилитися до вибору SARIMA/SARIMAX моделей.

Далі розглянемо деяку динаміку метрик, необхідних для розрахунку навантаження на КЦ за допомогою формули Ерланга. Для прикладу візьмемо АНТ – дуже важливий показник для розрахунку навантаження (рисунок 3.5 - 3.7).



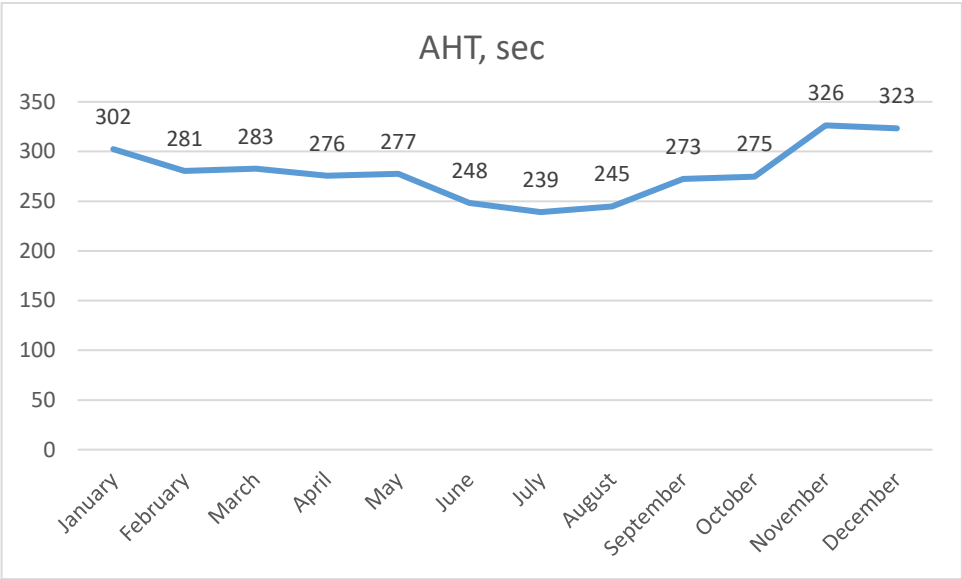


Рисунок 3.5 – Щомісячна динаміка середнього часу обробки одного  
вхідного дзвінка

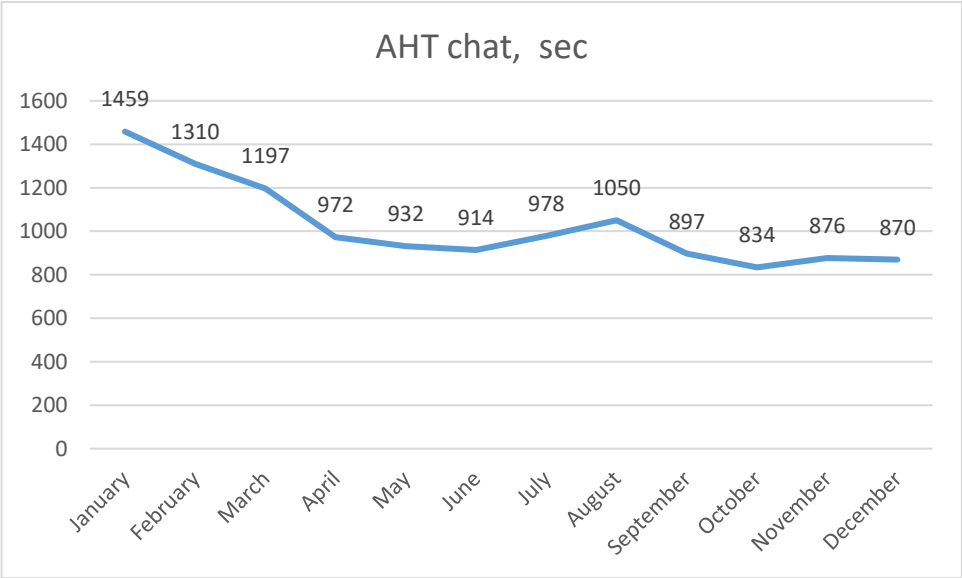


Рисунок 3.6 – Щомісячна динаміка середнього часу обробки одного  
чату

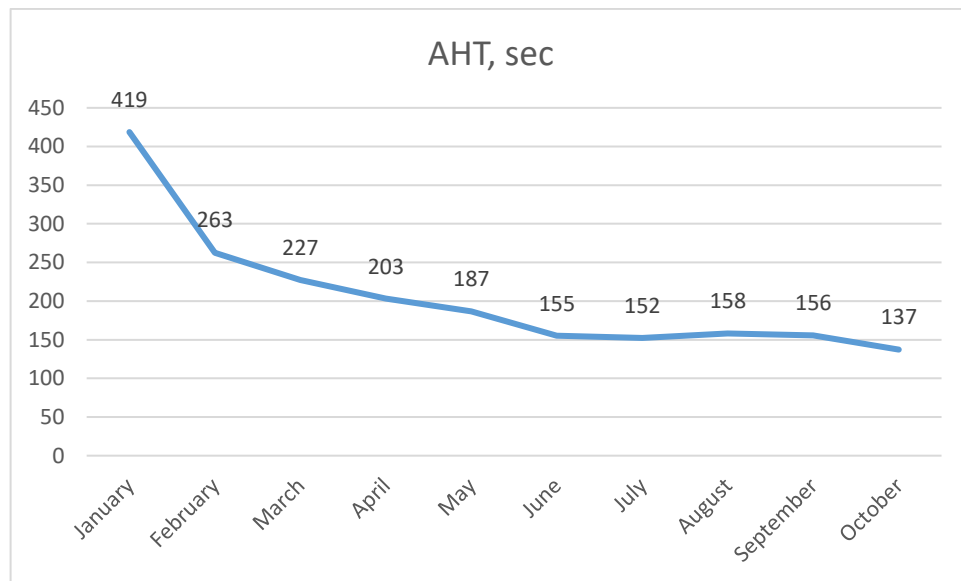


Рисунок 3.7 – Щомісячна динаміка середнього часу обробки одного електронного листа

### 3.3 Побудова моделей

#### 3.3.1 Тест на нормальність

Нормальний розподіл є найважливішим розподілом ймовірностей у статистиці, оскільки багато процесів у природі мають саме нормальний розподіл. Деякі приклади – це вік, зріст, вага та кров'яний тиск людини. Авторегресійні моделі передбачають, що помилки або залишки мають нормальний розподіл.

Отже нормальність даних – необхідна умова в нашому випадку.

Для перевірки даних на нормальність обрано тест Шапіро-Вілкса. Це найпотужніший тест для перевірки нормальності змінної. Його запропонували в 1965 році Самуель Санфорд Шапіро та Мартін Вілкс.

Оригінальна статистика тесту Шапіро-Вілка визначається наступним чином:

$$W = \frac{(\sum_{i=1}^m a_i y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

де  $y_i$  – статистика  $i$ -го порядку,

$\bar{y}$  – середнє значення вибірки,

$$\vec{a}_i = (a_1, \dots, a_n) = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{\frac{1}{2}}},$$

$\vec{m} = (m_1, \dots, m_n)^T$  – очікувані значення статистики порядку незалежних та однаково розподілених випадкових величин, відібраних із стандартної нормального розподілу;

$V$  – матриця коваріації статистики цих порядків.

Значення  $W$  лежить між нулем та одиницею. Невеликі значення  $W$  призводять до відхилення нормальності, тоді як значення одиниці вказує на нормальність даних [12].

Інтерпретація результатів тесту:

- якщо  $p$ -значення  $\leq 0,05$ , тоді ми відхиляємо нульову гіпотезу, тобто припускаємо, що розподіл нашої змінної не є нормальним;
- якщо  $p$ -значення  $> 0,05$ , тоді ми не можемо відкинути нульову гіпотезу, тобто ми вважаємо, що розподіл нашої змінної є нормальним [13].

Подивимося на результати тесту для оригінальних даних (таблиця 3.1):

Таблиця 3.1 – Результати тесту Шапіро-Вілкса на оригінальних даних

Черга	Дзвінки	Чати	Електронні листи
DE	0.36	0.01	0.72
FR	0.00	0.00	0.03
EN	0.00	0.29	0.43
ES	0.00	0.07	0.97
IT	0.00	0.21	0.13
PT	0.00	0.31	0.07
BG	0.00	0.16	0.01
NL	0.00	0.00	0.88
BN	0.05	0.01	0.99
SE	0.00	0.28	0.50
CZ	0.00	0.00	0.86
DK	0.00	0.24	0.88
FI	0.00	0.01	0.06
GR	0.00	0.01	0.01
HU	0.00	0.00	0.12
LT+LV	0.00	0.00	0.92
EE	0.00	0.00	0.00
NO	0.00	0.10	0.64
PL	0.00	0.04	0.50
RO	0.68	0.13	0.15
SK	0.00	0.01	0.26
SL	0.00	0.01	0.00

Можна помітити, що дані в основному не розподілені нормально, отже, необхідно застосувати деякі перетворення.

Оскільки маємо додатні значення, для нормалізації даних я вирішила використати звичайну функцію логарифму:

$$y = \ln(x)$$

Подивимося на результати тесту після перетворення (таблиця 3.2):

Таблиця 3.2 – Результати тесту Шапіро-Вілка після перетворення  
(3.2)

Черга	Дзвінки	Чати	Електронні листи
DE	0.66	0.14	0.08
FR	0.08	0.24	0.11
EN	0.27	0.27	0.27
ES	0.72	0.25	0.06
IT	0.36	0.14	0.18
PT	0.12	0.13	0.32
BG	0.39	0.33	0.23
NL	0.16	0.32	0.25
BN	0.07	0.08	0.08
SE	0.09	0.22	0.24
CZ	0.17	0.03	0.18
DK	0.15	0.05	0.25
FI	0.10	0.18	0.10
GR	0.07	0.10	0.08
HU	0.13	0.27	0.08
LT+LV	0.14	0.15	0.30
EE	0.39	0.96	0.15
NO	0.47	0.29	0.14
PL	0.11	0.04	0.45
RO	0.30	0.35	0.11
SK	0.44	0.06	0.22
SL	0.79	0.05	0.07

### 3.3.2 Вибір параметрів для моделі

Розглянемо SARIMAX модель, оскільки при підборі параметрів, ми можемо прийти до будь-якої моделі авторегресії, що описано в розділі 2.3.

Оскільки часових рядів занадто багато і підбір параметрів «на око» (за допомогою ACF та PACF) є трудомістким я написала алгоритм, який за допомогою мінімізації значення критерію Акаїке шляхом підбору всіх можливих комбінацій з обмеженнями на параметри  $p$  та  $q$  знаходить найкращу пару параметрів (несезонної та сезонної складових) з усіх можливих.

Отже маємо наступні результати (рисунок 3.8 – 3.9):

Calls	Chats	Mails
(1, 1, 1),	(0, 1, 1),	(0, 2, 0),
(2, 1, 1),	(0, 1, 0),	(0, 1, 1),
(1, 1, 0),	(0, 1, 0),	(1, 1, 2),
(0, 1, 0),	(0, 1, 0),	(0, 1, 1),
(0, 1, 1),	(0, 1, 1),	(0, 1, 1),
(0, 1, 0),	(0, 1, 0),	(0, 1, 1),
(2, 1, 2),	(2, 1, 0),	(0, 1, 1),
(1, 0, 2),	(0, 1, 0),	(0, 1, 1),
(2, 1, 1),	(0, 1, 0),	(0, 1, 1),
(0, 1, 0),	(0, 1, 0),	(0, 1, 0),
(0, 1, 0),	(0, 1, 0),	(0, 1, 0),
(0, 1, 0),	(0, 1, 0),	(0, 1, 1),
(0, 1, 0),	(1, 1, 1),	(0, 1, 2),
(0, 1, 0),	(0, 1, 0),	(0, 1, 1),
(0, 0, 0),	(0, 1, 0),	(0, 1, 0),
(0, 1, 0),	(0, 1, 0),	(0, 1, 2),
(0, 1, 0),	(0, 1, 1),	(0, 2, 0),
(0, 1, 0),	(0, 1, 0),	(0, 1, 1),
(1, 1, 1),	(0, 1, 0),	(0, 1, 1),
(0, 1, 1),	(0, 1, 0),	(0, 1, 0),
(0, 1, 0),	(0, 1, 2),	(0, 1, 1),
(0, 0, 0),	(0, 1, 0),	(0, 1, 1),

Рисунок 3.8 – Параметри несезонної частини моделі

Calls	Chats	Mails
[(2, 0, 0, 12),	[(1, 0, 0, 12),	[(0, 2, 0, 12),
(1, 0, 0, 12),	(0, 1, 0, 12),	(0, 1, 0, 12),
(2, 0, 0, 12),	(0, 1, 0, 12),	(1, 0, 0, 12),
(1, 0, 0, 12),	(1, 0, 0, 12),	(0, 1, 0, 12),
(1, 0, 0, 12),	(1, 0, 0, 12),	(0, 1, 0, 12),
(1, 0, 0, 12),	(1, 0, 0, 12),	(0, 1, 0, 12),
(2, 0, 0, 12),	(0, 1, 0, 12),	(2, 0, 0, 12),
(0, 1, 0, 12),	(1, 0, 0, 12),	(1, 0, 0, 12),
(0, 1, 0, 12),	(0, 1, 0, 12),	(1, 0, 0, 12),
(1, 0, 0, 12),	(1, 0, 0, 12),	(1, 0, 0, 12),
(2, 0, 0, 12),	(1, 2, 0, 12),	(1, 0, 0, 12),
(1, 0, 0, 12),	(1, 0, 0, 12),	(0, 0, 1, 12),
(2, 0, 0, 12),	(1, 0, 0, 12),	(0, 1, 0, 12),
(1, 0, 0, 12),	(1, 0, 0, 12),	(0, 1, 0, 12),
(0, 1, 0, 12),	(0, 1, 0, 12),	(1, 0, 0, 12),
(1, 0, 0, 12),	(0, 1, 0, 12),	(1, 0, 0, 12),
(1, 2, 0, 12),	(1, 0, 0, 12),	(0, 2, 0, 12),
(0, 1, 0, 12),	(0, 1, 0, 12),	(0, 1, 0, 12),
(2, 0, 0, 12),	(0, 1, 0, 12),	(2, 0, 0, 12),
(0, 1, 0, 12),	(1, 0, 0, 12),	(2, 0, 0, 12),
(1, 0, 0, 12),	(1, 2, 0, 12),	(1, 0, 0, 12),
(0, 1, 0, 12),	(1, 0, 0, 12),	(0, 1, 0, 12),

Рисунок 3.9 – Параметри сезонної частини моделі

Отже, після підбору параметрів моделі, настав час застосувати їх для прогнозу.

### 3.3.3 Прогноз

Як прийнято, ділимо відомі дані на train і test частини (дані, на яких тренується модель, а також відомі дані, завдяки яким перевіряється точність моделі) у співвідношенні 80% до 20 %.

На наступних зображеннях можна побачити деякі результати прогнозу (рисунок 3.10 – 3.12):

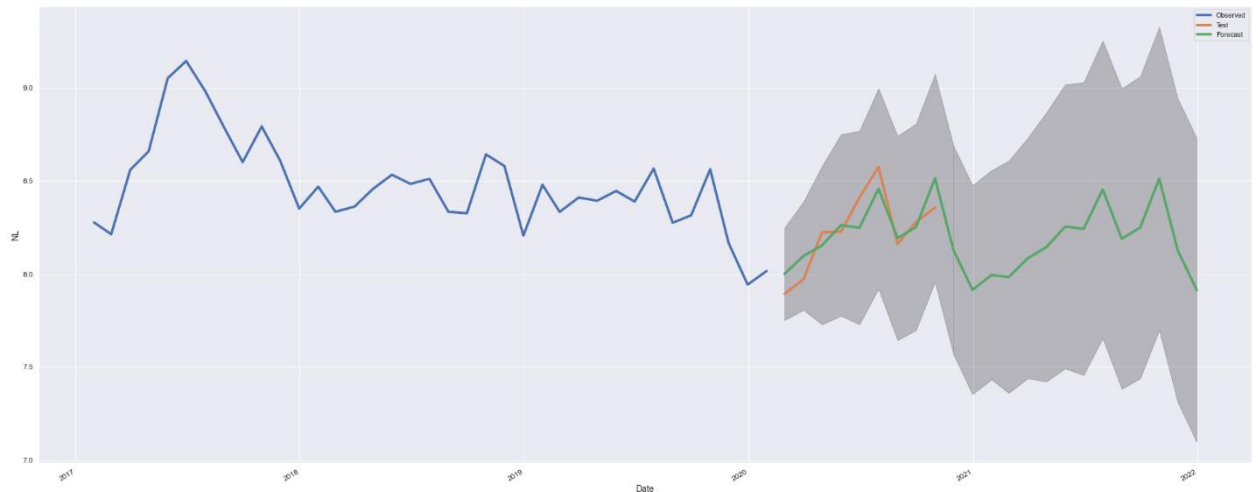


Рисунок 3.10 – Результат прогнозу вхідних дзвінків для одного з каналів

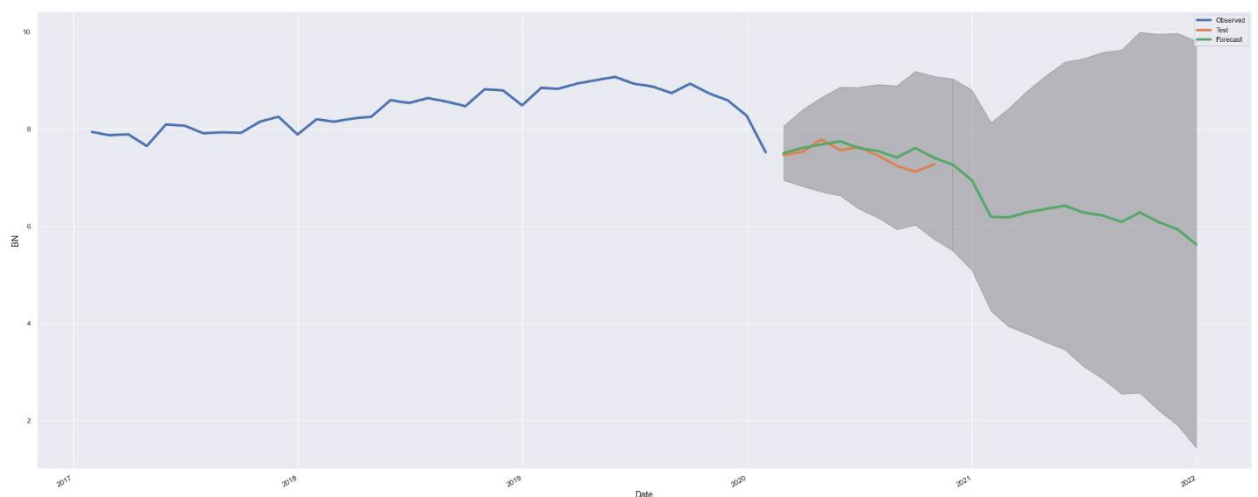


Рисунок 3.11 – Результат прогнозу чатів для одного з каналів



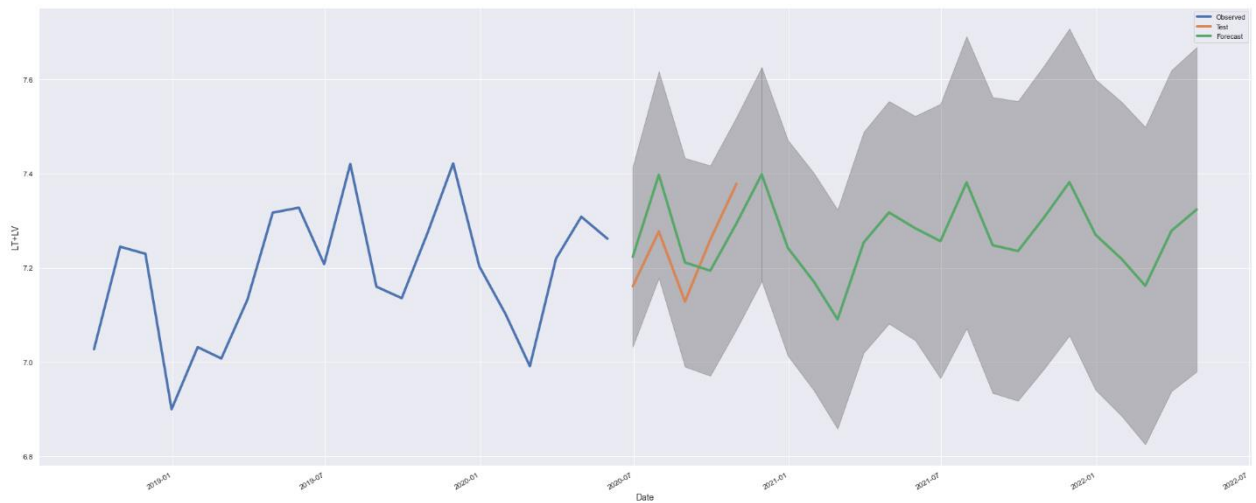


Рисунок 3.12 – Результат прогнозу чаті для одного з каналів

З рисунків вище можна зазначити, що з урахуванням доволі невеликої доступності історії і неможливості детально проаналізувати вхідні данні, ми отримали доволі предикативний на вигляд прогноз [14].

Настав час перевірити це за допомогою метрик.

### 3.3.4 Аналіз результатів прогнозу

#### 1) MAPE

Розраховується за формулою:

$$MAPE = \frac{1}{n} \sum_{i=1}^N \left| \frac{A_i - F_i}{A_i} \right|,$$

де  $n$  – довжина порівнюваних даних,

$A_i$  – справжні значення,

$F_i$  – прогнозовані значення.

Результати цієї метрики можна інтерпретувати за допомогою наступної таблиці (таблиця 3.3) [15]:

Таблиця 3.3 – Інтерпретація результатів прогнозу за допомогою метрики MAPE

Значення MAPE	Точність прогнозу
$\leq 10 \%$	Дуже точний
11 – 20 %	Точний
21 – 50 %	Значимий
$\geq 51 \%$	Неточний

Отже, подивимося на реальні результати (таблиця 3.4):

Таблиця 3.4 – Значення MAPE

	Calls	Chats	Mails
DE	2%	4%	76%
FR	2%	3%	4%
EN	4%	4%	2%
ES	3%	3%	0%
IT	3%	1%	2%
PT	4%	1%	2%
BG	3%	6%	1%
NL	2%	2%	2%
BN	12%	3%	2%
SE	7%	3%	2%
CZ	2%	42%	2%
DK	4%	2%	2%
FI	7%	2%	43%
GR	2%	4%	5%
HU	21%	19%	4%
LT+LV	7%	19%	2%
EE	11%	7%	84%
NO	16%	13%	5%
PL	2%	7%	1%
RO	4%	2%	4%
SK	8%	40%	3%
SL	80%	5%	5%

Отже, у більшості випадків маємо доволі хороший результат.

## 2) RMSE

Розраховується за формулою [16]:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (A_i - F_i)^2}{N}},$$

де  $n$  – довжина порівнюваних даних,

$A_i$  – справжні значення,

$F_i$  – прогнозовані значення.

Аналогічно до MAPE, чим краще RMSE, тим точніше значення прогнозу.

Отже, маємо наступні значення RMSE (таблиця 3.5):

Таблиця 3.5 – Значення RMSE

	Calls	Chats	Mails
DE	0.23	0.42	0.16
FR	0.33	0.20	0.18
EN	0.56	0.28	0.13
ES	0.36	0.13	0.10
IT	0.54	0.20	0.12
PT	0.40	0.20	0.11
BG	0.69	0.13	0.13
NL	0.52	0.29	0.24
BN	0.45	0.20	0.14
SE	0.59	0.25	0.19
CZ	0.21	0.97	0.23
DK	0.39	0.24	0.13
FI	0.76	0.18	0.10
GR	0.22	0.56	0.12
HU	0.78	0.29	0.31
LT+LV	0.41	0.22	0.13
EE	0.54	0.34	0.10
NO	1.22	0.56	0.18
PL	0.46	0.29	0.08
RO	0.38	0.23	0.49
SK	0.60	0.54	0.23
SL	0.69	0.33	0.09

Отже, результати RMSE теж є доволі задовільними, що дозволяє нам сказати, що результати прогнозу є достатньо точними і приймаються для подальшого їх застосування в розрахунках.

### 3.4 Розрахунок навантаження

Оскільки результати розрахунків є конфіденціальними і неінформативними загалом, в цьому розділі я опишу алгоритм, який я розробила для отримання цих розрахунків.

Нагадаю, що ціллю цих розрахунків є планування бюджету контактного центру на рік вперед. На виході ми отримуємо необхідну кількість людей щомісяця, що забезпечить бажаний рівень обслуговування для кожної черги. У цьому розділі часто застосовуватимуться метрики, описані у розділі 1.4.

Дані, необхідні на вхід:

- Прогнозовані дані у вигляді часових рядів.
- Бажане значення Service Level
- Shrinkage – розрахований з урахуванням відпусток, лікарняних та святкових днів.
- Середнє значення АНТ для кожного виду взаємодії за останній місяць.
- Середнє значення Abandon Time дзвінків та чатів за останній місяць.
- Необхідно знати щоденний розклад, оскільки в цьому КЦ агенти не працюють цілодобово.
- Розподіл навантаження протягом дня.

Алгоритм:

– Знаючи середнє навантаження, а також середню кількість дзвінків у будній день, можемо розрахувати кількість дзвінків, що припадає на кожну робочу годину.

– Знаючи бажаний рівень обслуговування, кількість дзвінків на годину за допомогою формули Ерланга С розраховуємо необхідну кількість персоналу, необхідного, щоб за вказану годину обробити взаємодії з урахуванням Shrinkage.

### **3.5 Висновки до розділу**

В цьому розділі описана практична складова магістерської дисертації: розраховано прогноз, виміряна його точність та описаний алгоритм для його практичного застосування.

## **РОЗДІЛ 4. СТАРТАП-ПРОЕКТ «Contact Center Automation»**

### **4.1 Вступ**

Суть цієї магістерської дисертації полягає у розробці і програмній реалізації алгоритму для прогнозування навантаження на контактний центр. Оскільки наразі дуже багато контактних центрів розраховують цей показник застарілими ручними методами, на основі реалізованого мною алгоритму можна запустити стартап-проект «Contact Center Automation», що з англійської перекладається як «Автоматизація Контактного Центру».

### **4.2 Опис ідеї проекту**

У таблиці 4.1 надано зміст ідеї (що пропонується), можливі напрямки застосування та основні вигоди, що може отримати користувач товару.



Таблиця 4.1 – Опис ідеї стартап-проекту

Зміст ідеї	Напрямки застосування	Вигоди для користувача
Створення програми, яка за допомогою сучасних методів прогнозування автоматизовано розраховуватиме навантаження контактного центру в розрізах, яких хоче бачити замовник	1. Якісне прогнозування вхідних взаємодій	1. Не потрібно залучати, або наймати працівників компанії для цієї задачі, оскільки можна придбати готовий продукт
	2. Розрахунок навантаження сучасними методами	2. Не завжди наймані працівники КЦ вміють реалізовувати програмні продукти рівня вище Excel. В даному випадку все це вже буде реалізовано на Python і підтверджено математично
	3. Можливість отримувати результати в різних розрізах (наприклад, каналів зв'язку: дзвінки, електронні листи тощо)	3. При створенні нових можливостей контакт-центру, їх тем можна буде додати в аналіз

Виділимо такі техніко-економічні характеристики ідеї:

- 1) якісний програмний продукт;
- 2) підтримка продукту протягом всієї експлуатації;
- 3) багатофункціональний аналіз і точний прогноз.

Для порівняння цього проекту з іншими представниками на ринку, у якості конкурентів виберемо такий програмний продукт як калькулятор Ерланга (на просторах інтернету їх дуже багато, але всі вони виконують одну і ту ж функцію без явних переваг (таблиця 4.2)).

Таблиця 4.2 – Визначення сильних, слабких та нейтральних характеристик ідеї проекту

	Техніко-економічні характеристики ідеї	(потенційні) товари/концепції конкурентів	W (слабка сторона)	N (нейтральна сторона)	S (сильна сторона)
		Калькулятор Ерланга			
	Якісний програмний продукт	+			+
	Підтримка продукту протягом всієї експлуатації	-		+	
	Багатофункціональний аналіз і точний прогноз	-	+		

### 4.3 Технологічний аудит ідеї проекту

Визначення технологічної здійсненності ідеї проекту передбачає аналіз складових, наведених у таблиці 4.3.

Таблиця 4.3 – Технологічна здійсненність ідеї проекту

п/п	Ідея проекту	Технології її реалізації	Наявність технологій	Доступність технологій
	Оренда офісу	Вибір зручного розташування	+	+
	Закупівля програмного забезпечення	Підбір програмного забезпечення, яке забезпечить всі потреби в реалізації і підтримці програми	+	+
	Закупівля технічного обладнання для реалізації	Закупівля ноутбуків, екранів та іншого обладнання для комфортної роботи всіх видів персоналу	+	+
Обрана технологія реалізації ідеї проекту: офіс з доступним програмним і технічним забезпеченням.				

#### 4.4 Аналіз ринкових можливостей запуску стартап-проекту

Спочатку проводиться аналіз попиту: наявність попиту, обсяг, динаміка розвитку ринку (таблиця 4.4).

Таблиця 4.4 – Попередня характеристика потенційного ринку стартап-проекту

п/п	Показники стану ринку (найменування)	Характеристика
	Кількість головних гравців, од	7
	Загальний обсяг продаж, грн	10 млн грн
	Динаміка ринку (якісна оцінка)	Зростає
	Наявність обмежень для входу	Отримання патенту та ліцензій
	Специфічні вимоги до стандартизації та сертифікації	Немає
	Середня норма рентабельності в галузі (або по ринку), %	25

Надалі визначаються потенційні групи клієнтів, їх характеристики, та формується орієнтовний перелік вимог до товару для кожної групи (таблиця 4.5).

Таблиця 4.5 – Характеристика потенційних клієнтів стартап-проекту

п/п	Потреба, що формує ринок	Цільова аудиторія (цільові сегменти ринку)	Відмінності у поведінці різних потенційних цільових груп клієнтів	Вимоги споживачів до товару
	Покупка реалізованої програми	Контактні центри компаній різноманітного спрямування	Різні масштаби (кількість працівників) КЦ, доступність історичних даних	Зручність інтерфейсу програми для використання працівників різного рівня технічної підготовки

Після визначення потенційних груп клієнтів проводиться аналіз ринкового середовища: складаються таблиці факторів, що сприяють ринковому впровадженню проекту, та факторів, що йому перешкоджають (таблиці 4.6 і 4.7). Фактори в таблиці подані в порядку зменшення значущості.

Таблиця 4.6 – Фактори загроз

п/п	Фактор	Зміст загрози	Можлива реакція компанії
	Погана якість даних зі сторони контактного центру	Кожен КЦ має своє програмне забезпечення для збору даних. Якість деяких із них може бути сумнівною.	Запропонувати перейти на інше програмне забезпечення, в якості якого ми впевнені.
	Можливий людський фактор (помилки в програмній реалізації)	В кожній програмі завжди є свої помилки, що зазвичай викликає незадоволення зі сторони клієнта.	Запропонувати знижку для того, щоб клієнт не відмовлявся від підписки, або не розривав договір
	Недостатня кількість історичних даних для точного прогнозу	Маленька кількість історичних даних може вплинути а якість прогнозованих даних	Попередити про це клієнта і запропонувати відкласти впровадження цієї компанії

Таблиця 4.7 – Фактори можливостей

п/п	Фактор	Зміст можливості	Можлива реакція компанії
	Додаткова можливість аналітичного супроводу	З часом можна розширювати можливості програмного забезпечення, наприклад давати аналітичний супровід отриманих результатів у вигляді графіків, висновків та порад	Залежить від бюджету і потреб контактного центру

Далі проводиться аналіз пропозиції: визначаються загальні риси конкуренції на ринку (таблиця 4.8).

Таблиця 4.8 – Ступеневий аналіз конкуренції на ринку

Особливості конкурентного середовища	В чому проявляється дана характеристика	Вплив на діяльність підприємства (можливі дії компанії, щоб бути конкурентоспроможною)
1. Вказати тип конкуренції - монополія / олігополія / монополістична / чиста	Чиста	Гарні перспективи розвитку
2. За рівнем конкурентної боротьби - локальний / національний / ...	Національний	Ведучи конкуренцію на національному рівні, компанії необхідно прикласти належні зусилля для охоплення всього національного ринку
3. За галузевою ознакою - міжгалузева / внутрішньогалузева	Внутрішньогалузева	Необхідно зосередити зусилля на пошуку конкурентних переваг, які дозволять компанії займати стійкі конкурентні позиції на даному ринку
4. Конкуренція за видами товарів: - товарно-родова - товарно-видова - між бажаннями	Товарно-родова	
5. За характером конкурентних переваг - цінова / нецінова	Нецінова	Перевагою є унікальність на ринку на даний момент
6. За інтенсивністю - марочна / не марочна	Марочна	

Після аналізу конкуренції проводиться більш детальний аналіз умов конкуренції в галузі (таблиця 4.9).

Таблиця 4.9 – Аналіз конкуренції в галузі за М. Портером

	Прямі конкуренти в галузі	Потенційні конкуренти	Постачальники	Клієнти	Товари-замінники
Складові аналізу	Навести перелік прямих конкурентів	Визначити бар'єри входження в ринок	Визначити фактори сили постачальників	Визначити фактори сили споживачів	Фактори загроз з боку замінників
Висновки:	Прямих конкурентів немає	Отримання патентів і ліцензій	Немає постачальників	Немає	Немає

За результатами аналізу таблиці робиться висновок щодо принципової можливості роботи на ринку з огляду на конкурентну ситуацію (таблиця 4.10).

Таблиця 4.10 – Обґрунтування факторів конкурентоспроможності

п/п	Фактор конкурентоспроможності	Обґрунтування (наведення чинників, що роблять фактор для порівняння конкурентних проектів значущим)
	Відносна унікальність на ринку	Єдина конкуренція на ринку – низка безкоштовних негнучких програм-калькуляторів в інтернеті
	Математично обґрунтовані підходи	Перевагою є те, що програмна реалізація є повністю математично обґрунтованою
	Повна підтримка програмного забезпечення після впровадження	Після того, як продається програмне забезпечення, воно продовжує покращуватися, виправлятися помилки і т.д.



За визначеними факторами конкурентоспроможності (таблиця 4.10) проводиться аналіз сильних та слабких сторін стартап-проекту (таблиця 4.11).

Таблиця 4.11 – Порівняльний аналіз сильних та слабких сторін «Contact Center Automation»

п/п	Фактор конкурентоспроможності	Бали 1-20	Рейтинг товарів-конкурентів у порівнянні з «Contact Center Automation»						
			-3	-2	-1		1	2	3
	Відносна унікальність на ринку	20	+						
	Математично обґрунтовані підходи	10					+		
	Повна підтримка програмного забезпечення після впровадження	17		+					

Фінальним етапом ринкового аналізу можливостей впровадження проекту є складання SWOT-аналізу (матриці аналізу сильних (Strength) та слабких (Weak) сторін, загроз (Troubles) та можливостей (Opportunities) (таблиця 4.12).

Таблиця 4.12 – SWOT- аналіз стартап-проекту

Сильні сторони: – Відсутність реальних конкурентів – Гарантія коректних результатів і математичне обґрунтування	Слабкі сторони: - Програмне забезпечення є платним і не кожна компанія хоче витратити на це гроші - Відсутність аналогічного досвіду
Можливості: – Впровадження нових реалізацій в програмі	Загрози: - Поява реального конкурента на ринку

На основі SWOT-аналізу розробляються альтернативи ринкової поведінки (перелік заходів) для виведення стартап-проекту на ринок та орієнтовний оптимальний час їх ринкової реалізації з огляду на потенційні проекти конкурентів, що можуть бути виведені на ринок.

Визначені альтернативи аналізуються з точки зору строків та ймовірності отримання ресурсів (таблиця 4.13).

Таблиця 4.13 – Альтернативи ринкового впровадження стартап-проекту

п/п	Альтернатива (орієнтовний комплекс заходів) ринкової поведінки	Ймовірність отримання ресурсів	Строки реалізації
	Диверсифікація	Створення додаткових можливостей програми, наприклад, супроводжуюча аналітика	6 місяців
	Розвиток ринку	Вихід продукту на міжнародний ринок	2 роки

#### 4.5 Розроблення ринкової стратегії проекту

Розроблення ринкової стратегії першим кроком передбачає визначення стратегії охоплення ринку: опис цільових груп потенційних споживачів (таблиця 4.14).

Таблиця 4.14 – Вибір цільових груп потенційних споживачів

п/п	Опис профілю цільової групи потенційних клієнтів	Гото вність споживачів сприйняти продукт	Орієнт овний попит в межах цільової групи (сегменту)	Інтенсив ність конкуренції в сегменті	Важ кість входу у сегмент
	Контактний центр	Висока	Високий	Низька	Середня
Які цільові групи обрано: Контактний центр					

Для роботи в обраних сегментах ринку необхідно сформувати базову стратегію розвитку (таблиця 4.15).

Таблиця 4.15 – Визначення базової стратегії розвитку

п/п	Обрана альтернатива розвитку проекту	Стратегія охоплення ринку	Ключові конкурентоспроможні позиції відповідно до обраної альтернативи	Базова стратегія розвитку
	Спеціалізація	Зосередження на одному сегменті ринку	Науковість підходу, гарантії ефективної роботи	Стратегія спеціалізації

Наступним кроком є вибір стратегії конкурентної поведінки (таблиця 4.16).

Таблиця 4.16 – Визначення базової стратегії конкурентної поведінки

п/п	Чи є проект «першопроходцем» на ринку?	Чи буде компанія шукати нових споживачів, або забирати існуючих у конкурентів?	Чи буде компанія копіювати основні характеристики товару конкурента, і які?	Стратегія конкурентної поведінки
	Так	Так	Ні	Стратегія лідера

На основі вимог споживачів з обраних сегментів до постачальника (стартап-компанії) та до продукту, а також в залежності від обраної базової стратегії розвитку (таблиця 4.15) та стратегії конкурентної поведінки (таблиця 4.16) розробляється стратегія позиціонування (таблиця 4.17). що полягає у формуванні ринкової позиції (комплексу асоціацій), за яким споживачі мають ідентифікувати проект.

Таблиця 4.17 – Визначення стратегії позиціонування

п/п	Вимоги до товару цільової аудиторії	Базова стратегія розвитку	Ключові конкурентоспроможні позиції власного стартап-проекту	Вибір асоціацій, які мають сформувати комплексну позицію власного проекту (три ключових)
	Ефективність; відповідальність; новизна на ринку.	Стратегія спеціалізації	Науковість підходу, гарантії ефективної роботи, перспективи підтримки і подальшого розвитку	Науковість, точність, новизна

#### 4.6 Розроблення маркетингової програми стартап-проекту

Першим кроком є формування маркетингової концепції товару, який отримає споживач. Для цього у таблиці 4.18 потрібно підсумувати результати попереднього аналізу конкурентоспроможності товару.

Таблиця 4.18. Визначення ключових переваг концепції потенційного товару

п/п	Потреба	Вигода, яку пропонує товар	Ключові переваги перед конкурентами (існуючі або такі, що потрібно створити)
	Автоматизація процесів	Реалізація цієї автоматизації	У конкурентів данні необхідно вводити вручну, що дуже незручно для великих об'ємів даних
	Аналітичний супровід	Реалізація аналітичного а технічного супроводу	Такого немає у жодного з і так слабких конкурентів

Надалі розробляється трирівнева маркетингова модель товару: уточнюються ідея продукту та/або послуги, його фізичні складові, особливості процесу його надання (таблиця 4.19).

Таблиця 4.19 – Опис трьох рівнів моделі товару

Рівні товару	Сутність та складові		
I. Товар за задумом	Розумний програмний продукт із високим рівнем точності оцінювання та зручним інтерфейсом.		
II. Товар у реальному виконанні	Властивості/характеристики	М/Нм	Вр/Тх /Тл/Е/Ор
	1. Функціональність	М	Тх
	2. Швидкодія	М	Тх
	3. Зручність	М	Е
	4. Зовнішній вигляд інтерфейсу	М	Е/Ор
	Якість: продукт має відповідати міжнародним стандартам		
III. Товар із підкріпленням	Використовується реклама		
	Використовуються тимчасові знижки		
За рахунок чого потенційний товар буде захищено від копіювання: за рахунок електронних ключів та інтелектуальної власності.			

Наступним кроком є визначення цінових меж, якими необхідно керуватись при встановленні ціни на потенційний товар (остаточне визначення ціни відбувається під час фінансово-економічного аналізу проєкту), яке передбачає аналіз ціни на товари-аналоги або товари субститути, а також аналіз рівня доходів цільової групи споживачів (таблиця 4.20).

Таблиця 4.20 – Визначення меж встановлення ціни

п/п	Рівень цін на товари-замінники	Рівень цін на товари-аналоги	Рівень доходів цільової групи споживачів	Верхня та нижня межі встановлення ціни на товар/послугу
	Немає	Вищий	Високий	Індивідуальна домовленість в залежності від складності інтеграції

Наступним кроком є визначення оптимальної системи збуту, в межах якого приймається рішення (таблиця 4.21).

Таблиця 4.21 – Формування системи збуту

п/п	Специфіка закупівельної поведінки цільових клієнтів	Функції збуту, які має виконувати постачальник товару	Глибина каналу збуту	Оптимальна система збуту
	Компанія надає свої анонімізовані данні в користування	Результати і аналітичний супровід		

Останньою складовою маркетингової програми є розроблення концепції маркетингових комунікацій, що спирається на попередньо обрану основу для позиціонування, визначену специфіку поведінки клієнтів (таблиця 4.22).

Таблиця 4.22 – Концепція маркетингових комунікацій

п/п	Специфіка поведінки цільових клієнтів	Канали комунікацій, якими користуються цільові клієнти	Ключові позиції, обрані для позиціонування	Завдання рекламного повідомлення	Концепція рекламного звернення
	Вимагають інтерпретацію результатів і швидкого реагування у випадку помилок	Телефон, електронна пошта, індивідуальні планові зустрічі	Гарантія результату, швидка підтримка, знижки для постійних клієнтів	Привернути увагу та зацікавити цільових клієнтів.	Надання знижок на перші місяці користування продуктом, а також для постійних клієнтів

#### **4.7 Висновки до розділу**

У результаті опрацювання цього розділу був розроблений комерційно спроможний проект «Contact Center Automation». Для цього були оглянуті та запропоновані ринкові стратегії; були описані потенційні конкуренти та переваги і недоліки у порівнянні з ними; були розглянуті бар'єри входження на ринок. Даний проект доцільно розвивати та реалізовувати.



## ВИСНОВКИ

В даній магістерській дисертації було реалізовано прогноз вхідних взаємодій контактного центру міжнародної компанії з продажу автомобільних запчастин. Після чого описаний алгоритм розрахунку навантаження на контактний центр з ціллю розрахунку необхідної кількості персоналу для успішної роботи контактного центру.

Головною складністю в даній роботі є масштабність КЦ та інтернаціональність компанії, через що потрібно було зробити прогнози на основі 66 часових рядів. Ця складність вплинула на неможливість детального аналізу кожного часового ряду і необхідність повної автоматизації в підготовці даних до прогнозу, підбору параметрів моделей, а також аналізу результатів прогнозів.

Особливістю моєї роботи є те, що ця задача була розроблена для реальної компанії і вже успішно використовується.

Новизною в моїй роботі є ідея застосування формули Ерланга, як інструменту для розрахунку навантаження на великій кількості часових рядів, оснований на прогнозі, а не реальних значеннях.

## ПЕРЕЛІК ПОСИЛАНЬ

1. What is a Contact Center? URL: <https://aws.amazon.com/ru/what-is-a-contact-center/> (Last accessed: 27.11.2020).
2. The Best Channels for Customer Service Within the Contact Center URL: <https://www.westuc.com/en-us/blog/managed-voice-services/best-channels-customer-service-within-contact-center> (Last accessed: 27.11.2020).
3. Luis Felipe Llanos Reynoso. A study on call/contact centers' inbound and outbound management process in Mexico. *South Asian Journal of Management Sciences*. Fall 2016. Vol. 10, No. 2. P. 2-3.
4. Call Center KPIs and Metrics 2020 [ULTIMATE GUIDE]. URL: <https://voiptimecloud.com/blog/take-charge-of-your-data-7-key-contact-center-kpis-explained> (Last accessed: 27.11.2020).
5. Rob J Hyndman and George Athanasopoulos. Forecasting: Principles and Practice. Monash University, Australia, 2018. 255 p.
6. SARIMAX: Introduction. URL: [https://www.statsmodels.org/stable/examples/notebooks/generated/spacespace\\_sarimax\\_stata.html](https://www.statsmodels.org/stable/examples/notebooks/generated/spacespace_sarimax_stata.html) (Last accessed: 27.11.2020).
7. Erlang Distribution: Definition, Examples. URL: <https://www.statisticshowto.com/erlang-distribution/> (Last accessed: 27.11.2020).
8. A Beginner's Guide to the Erlang A Formula. URL: <https://www.callcentrehelper.com/a-beginners-guide-to-the-erlang-a-formula-140998.htm> (Last accessed: 27.11.2020).
9. Lawrence Brown, Noah Gans, Avishai Mandelbaum. Statistical Analysis of a Telephone Call Center: A Queueing-Science Perspective. *Journal of the American Statistical Association*. March 2005. P. 59.

10. Erik Chromy, Tibor Misuth, Adam Weber. Application of Erlang Formulae in Next Generation Networks. *I. J. Computer Network and Information Security*. February 2012. Vol. 2, No. 2. P.59-66.
11. Nornadiah Mohd Razali, Bee Wah Yap. Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests. *Journal of Statistical Modeling and Analytic*. 2011. Vol.2, No.1. P. 21-33.
12. A. Antipov and N. Meade. Forecasting Call Frequency at a Financial Services Call Centre. *The Journal of the Operational Research Society*. 2002. Vol. 2, No. 3. P. 953–960.
13. Chompoonoot Kasemset1, Nisachon Sae-Haew, Apichat Sopadang. Multiple Regression Model for Forecasting Quantity of Supply of Off-season Longan. *Chiang Mai University Journal of Natural Sciences*. 2014. Vol. 13(3). P.13.
14. RMSE: Root Mean Square Error. URL: <https://www.statisticshowto.com/probability-and-statistics/regression-analysis/rmse-root-mean-square-error/> (Last accessed: 27.11.2020).

## ДОДАТОК А

### КОД ПРОГРАМНОГО ПРОДУКТУ

```

## Data preparation

import pyodbc
import pandas as pd

#### All inbound interactions since 01/11/2019 from Genesys, monthly data

query = open('Inbound_interactions.sql','r').read()

Inbound_interactions = pd.read_sql(query,
connection).sort_values(by=['Channel', 'Date', 'Country'])
Inbound_interactions

#### Splitting interactions into separate channels

Calls = Inbound_interactions[Inbound_interactions.Channel ==
'Calls'].drop(['Channel'], axis = 1)
Chats = Inbound_interactions[Inbound_interactions.Channel ==
'Chats'].drop(['Channel'], axis = 1)
Mails = Inbound_interactions[Inbound_interactions.Channel == 'E-
mails'].drop(['Channel'], axis = 1)

#### Making separate time series for each country and each channel

Calls = Calls.pivot(index='Date', columns='Country', values=
'Quantity').fillna(0).astype(int)
Calls.index = pd.to_datetime(Calls.index).date
Mails = Mails.pivot(index='Date', columns='Country', values=
'Quantity').fillna(0).astype(int)
Mails.index = pd.to_datetime(Mails.index).date
Chats = Chats.pivot(index='Date', columns='Country', values=
'Quantity').fillna(0).astype(int)
Chats.index = pd.to_datetime(Chats.index).date

#### Uploading old monthly data (before Genesis was integrated)
##### Calls and Chats - since 01/01/2017
##### Mails - since 01/09/2018

Calls_old = pd.read_csv('Calls_old data.csv').set_index('Date')
Calls_old.index = pd.to_datetime(Calls_old.index).date

Chats_old = pd.read_csv('Chats_old data.csv').set_index('Date')
Chats_old.index = pd.to_datetime(Chats_old.index).date

Mails_old = pd.read_csv('Mails_old data.csv').set_index('Date')
Mails_old.index = pd.to_datetime(Mails_old.index).date

#### Appending new data to the old data

Calls = Calls_old.append(Calls).fillna(0).astype(int)

```

```

Chats = Chats_old.append(Chats).fillna(0).astype(int)
Mails = Mails_old.append(Mails).fillna(0).astype(int)

s = np.zeros(len(Mails))

for i in Mails.columns:
    s += Mails[i].array
s

sums = pd.DataFrame(s)
sums.index = Mails.index

sums.plot(marker = 'o', linewidth = 4)
plt.yticks(fontsize = 21)
plt.xticks(np.arange(Mails.index[0],Mails.index[-1],len(Mails)),rotation =
90, fontsize = 18)
plt.title('Щомісячна динаміка кількості вхідних електронних листів', fontsize
= 30)
plt.show()

## Data analysis

import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np

Tables = [Calls,Chats,Mails]
Table_names = ['Calls','Chats','Mails']

sns.set(rc={'figure.figsize':(35, 15)})
k = 0
for i in Tables:
    print(Table_names[k])
    i[i.columns[:7]].plot(marker = 'o')
    plt.xticks(np.arange(i.index[0],i.index[-1],len(i)))
    plt.show()
    i[i.columns[7:15]].plot(marker = 'o')
    plt.xticks(np.arange(i.index[0],i.index[-1],len(i)))
    plt.show()
    i[i.columns[15:]].plot(marker = 'o')
    plt.xticks(np.arange(i.index[0],i.index[-1],len(i)))
    plt.show()
    k += 1

### Calls

sns.set(rc={'figure.figsize':(35, 15)})
Calls[Calls.columns[:7]].plot(marker = 'o',fontsize = 23)
plt.xticks(np.arange(Calls.index[0],Calls.index[-1],len(Calls)),rotation =
90, fontsize = 18)
plt.show()

Calls[Calls.columns[7:15]].plot(marker = 'o')
plt.xticks(np.arange(Calls.index[0],Calls.index[-1],len(Calls)))
plt.show()

Calls[Calls.columns[15:]].plot(marker = 'o')
plt.xticks(np.arange(Calls.index[0],Calls.index[-1],len(Calls)))
plt.show()

```

```

### Chats

Chats[Chats.columns[:7]].plot(marker = 'o')
plt.xticks(np.arange(Chats.index[0],Chats.index[-1],len(Chats)))
plt.show()

Chats[Chats.columns[7:15]].plot(marker = 'o')
plt.xticks(np.arange(Chats.index[0],Chats.index[-1],len(Chats)))
plt.show()

Chats[Chats.columns[15:]].plot(marker = 'o')
plt.xticks(np.arange(Chats.index[0],Chats.index[-1],len(Chats)))
plt.show()

### Mails

Mails[Mails.columns[:7]].plot(marker = 'o')
plt.xticks(np.arange(Mails.index[0],Mails.index[-1],len(Mails)))
plt.show()

Mails[Mails.columns[7:15]].plot(marker = 'o')
plt.xticks(np.arange(Mails.index[0],Mails.index[-1],len(Mails)))
plt.show()

Mails[Mails.columns[15:]].plot(marker = 'o')
plt.xticks(np.arange(Mails.index[0],Mails.index[-1],len(Mails)))
plt.show()

d = {'Country': Calls.columns, 'Calls': np.zeros(len(Calls.columns)),
      'Chats': np.zeros(len(Calls.columns)), 'Mails': np.zeros(len(Calls.columns))}
df = pd.DataFrame(data=d)
df.index = df['Country']
df = df.drop(columns = 'Country')

Calls = Calls[:-1]
Mails = Mails[:-1]
Chats = Chats[:-1]

Tables = [Calls,Mails,Chats]

### Checking time series for normality with a help of Shapiro-Wilk test

##### p - value > 0.05 means normality of the time-series

from scipy.stats import shapiro

def Shapiro_Test(Tables,table):
    k = 0
    for i in Tables:
        for j in i.columns:
            table[Table_names[k]][j] = shapiro(i[j])[1]
        k += 1
    return table

sh_test = df

Shapiro_Test(Tables_1,sh_test)

```

[illegible]

```

                                                    enforce_invertibility=False)
        results = mod.fit()
        if aic_prev == 0:
            best_param[j][i] = param
            best_param_seasonal[j][i] = param_seasonal
            aic_prev = results.aic
        elif results.aic < aic_prev:
            best_param[j][i] = param
            best_param_seasonal[j][i] = param_seasonal
            aic_prev = results.aic
    except:
        continue

import math

def RMSE(real,predict):
    return math.sqrt(sum(np.square(real-predict))/len(real))

def MAPE(real,predict):
    return sum(abs(real-predict)/real)/len(real)

def R_sq(real,predict):
    return 1 - sum(np.square(real - predict))/sum(np.square(real -
np.average(real)))

def get_prediction(train,test,forecast,p,predict_from):

    if p == 0:
        print('Calls')
    elif p == 1:
        print('Chats')
    else:
        print('Mails')

    for i in range(len(columns)):
        mod = sm.tsa.statespace.SARIMAX(train[columns[i]],
                                                    order=best_param[p][i],

seasonal_order=best_param_seasonal[p][i],

enforce_stationarity=True,

enforce_invertibility=False)
        results = mod.fit()
        pred = results.get_prediction(start=pd.to_datetime(predict_from),
dynamic=False)
        pred_ci = pred.conf_int()
        pred_uc = results.get_forecast(steps=forecast_length)
        forecast[columns[i]] = pred_uc.predicted_mean.to_numpy()
        pred_ciuc = pred_uc.conf_int()
        mape[p][i] =
MAPE(test[columns[i]].to_numpy(),pred.predicted_mean.to_numpy())
        r_sq[p][i] =
RMSE(test[columns[i]].to_numpy(),pred.predicted_mean.to_numpy())
        print('MAPE: ',mape[p][i])
        print('rmse: ',r_sq[p][i])
        ax = train[columns[i]].plot(label='Observed', linewidth = 4)
        test[columns[i]].plot(label='Test', linewidth = 4)
        #ax.title(columns[i])
        pred_uc.predicted_mean.plot(ax=ax, label='Forecast',linewidth = 4)

```



```

        ax.fill_between(pred_ci.index,
                        pred_ci.iloc[:, 0],
                        pred_ci.iloc[:, 1], color='k', alpha=.2)
        ax.fill_between(pred_ciuc.index,
                        pred_ciuc.iloc[:, 0],
                        pred_ciuc.iloc[:, 1], color='k', alpha=.25)
        ax.set_xlabel('Date')
        ax.set_ylabel(columns[i])
        plt.legend()
        plt.show()

for i in range(len(columns)):
    the_best_parameters(Calls_l,i,0)

for i in range(len(columns)):
    the_best_parameters(Chats_l,i,1)

for i in range(len(columns)):
    the_best_parameters(Mails_l,i,2)

for i in range(3):
    for j in range(22):
        if best_param_seasonal[i][j] == (0, 0, 0, 12):
            best_param_seasonal[i][j] = (0, 1, 0, 12)

for i in range(3):
    for j in range(22):
        if best_param[i][j] == (0, 0, 0):
            best_param[i][j] = (0, 1, 0)

forecast_calls = pd.DataFrame(columns = Train[0].columns)
forecast_chats = pd.DataFrame(columns = Train[1].columns)
forecast_mails = pd.DataFrame(columns = Train[2].columns)

forecast_length = 14 + test_length_calls_chats

print('*****Calls*****')
get_prediction(Train[0],Test[0],forecast_calls,0,'2020-11-30')

np.exp(forecast_calls.fillna(0))
.to_csv('C:/Users/User/Desktop/Budget/Budget/Forecast/Result/forecast_calls_n
ew.csv', sep = ';')

print('*****Chats*****')
get_prediction(Train[1],Test[1],forecast_chats,1,'2020-11-30')

np.exp(forecast_chats.fillna(0))
.to_csv('C:/Users/User/Desktop/Budget/Budget/Forecast/Result/forecast_chats_n
ew.csv', sep = ';')

print('*****Mails*****')
get_prediction(Train[2],Test[2],forecast_mails,2,'2020-11-30')

forecast_length = 14 + test_length_mails

np.exp(forecast_mails.fillna(0))
.to_csv('C:/Users/User/Desktop/Budget/Budget/Forecast/Result/forecast_mails_n
ew.csv', sep = ';')

mapes = pd.DataFrame(mape, columns = columns,

```

```
index=['Calls','Chats','Mails'])
mapes.to_csv('C:/Users/User/Desktop/Budget/Budget/Forecast/Result/MAPE_new.csv', sep = ';')

rmse = pd.DataFrame(r_sq, columns = columns, index=['Calls','Chats','Mails'])
rmse.to_csv('C:/Users/User/Desktop/Budget/Budget/Forecast/Result/RMSE.csv', sep = ';')

Calls.to_csv('C:/Users/User/Desktop/Budget/Budget/Forecast/Input/Calls.csv', sep = ';')

Chats.to_csv('C:/Users/User/Desktop/Budget/Budget/Forecast/Input/Chats.csv', sep = ';')

Mails.to_csv('C:/Users/User/Desktop/Budget/Budget/Forecast/Input/Mails.csv', sep = ';')
```